# Integrating Peer and Teacher Feedback in the Assessment of English as a Foreign Language Oral Presentations in Higher Education

*by* Polyglot: Journal of Linguistics, Literature, and Language Education

**Submission date:** 29-Dec-2025 09:36AM (UTC+0700)
**Submission ID:** 2849706500
**File name:** 54-69_76.docx (952.96K)
**Word count:** 5730
**Character count:** 33158

# Integrating Peer and Teacher Feedback in the Assessment of English as a Foreign Language Oral Presentations in Higher Education

**Montserrat Iglesias**
CETT-UB Barcelona School of Tourism, Hospitality and Gastronomy, University of Barcelona, Spain

montserrat.iglesias@cett.cat

**ABSTRACT**

This descriptive study investigates the integration of peer and teacher assessment in evaluating oral presentations among English as a Foreign Language (EFL) undergraduate students. The research followed a mixed-methods approach aiming to a) analyze the characteristics of peer feedback comments, and b) examine the alignment between peer and teacher scores to assess grading validity and reliability. Data were collected from 50 Spanish students over three academic years (2020-2021, 2023-2024, and 2024-2025), comprising 341 peer feedback forms that included both Likert-scale scores and open-ended comments focusing on four criteria: language use, task achievement, fluency and pronunciation, and communicative resources. Quantitative analyses were conducted to calculate means, standard deviations, Pearson correlation coefficients, and Cronbach's alpha values. Results indicated strong alignment between peer and teacher assessments in 2020-21 and 2024-25, but a moderate correlation and lower reliability in 2023-24, suggesting inflated peer scores. A qualitative exploration through content analysis and categorization revealed consistent feedback themes across years, with fluency, vocabulary, and pronunciation being the most frequently mentioned strengths and weaknesses. These findings suggest that, when supported by training and structured rubrics, peer assessment can be a valid and reliable complement to teacher assessment in EFL oral presentation grading.

Keywords: Assessment Criteria, Grading, Oral Presentations, Peer Feedback, Teacher Assessment

## INTRODUCTION

Although peer feedback and peer assessment have been increasingly explored in EFL (English as a Foreign Language) higher education contexts, the integration of these approaches specifically into the marking of oral presentations remains under-researched, particularly regarding the combination of peer and teacher assessment for summative purposes (Murillo-Zamorano & Montanero, 2018; Wanchid & Charoensuk, 2024). Most existing studies on co-evaluation in language learning focus on either formative peer feedback or comparing feedback types, but fewer investigate the impact of combining peer grading with teacher grading in high-stakes assessment (Maiz Arévalo, 2008; Yu, 2024). This research gap is significant in relation to oral presentation skills because they are crucial for academic and professional success, and effective assessment strategies can foster communicative competence, autonomy, and deeper engagement with learning outcomes (Prosenjak & Lučev, 2020; Wanchid & Charoensuk, 2024).

Practically, combining peer and teacher assessment can enhance student motivation and responsibility for learning (Prosenjak & Lučev, 2020; Wanchid & Charoensuk, 2024) while preparing students for real-world evaluative situations by exposing them to diverse perspectives (Maiz Arévalo, 2008; Wanchid & Charoensuk, 2024). This approach can also provide more comprehensive and balanced feedback (Iglesias, 2013; Liu & Aryadoust, 2024).

The study reported in this paper aimed at gaining insights into the integration of teacher and peer feedback in formative assessment processes within an EFL higher education context. This general objective was broken down into two specific research objectives. The first was to examine the nature of mobile-assisted peer feedback comments on students' oral presentations in EFL over three academic years, and derived from this research question: What are the main characteristics of mobile-assisted peer feedback comments on oral presentations? In turn, the second specific objective was to analyze the alignment of peer assessment and teacher assessment in grading EFL oral presentations, and stemmed from this question: Does the integration of peer and teacher assessment ensure the reliability and validity of oral presentation grading?

***Literature Review***

Socio-constructivism, as articulated by Vygotsky, posits that learning is a social process best supported through interaction and collaboration. Integrating peer feedback and assessment aligns with socio-constructivist theories, emphasizing collaborative knowledge construction, increased learner autonomy, and the development of metacognitive skills through active participation in assessment processes (Iglesias, 2013; Janesarvatan & Asoodar, 2024). Peer assessment enables students to co-construct knowledge, negotiate criteria, and internalize standards of quality through discussion and reflection (McGarrigle, 2013). This approach is reinforced by the concept of the zone of proximal development, where learners advance by receiving guidance from peers and teachers who possess varying levels of expertise (McGarrigle, 2013). Previous research has determined that both teachers and students view explicit feedback as important and effective, and that these beliefs are shaped by sociocultural and contextual aspects, as well as by personal experiences (Van Ha, 2021).

While teachers' feedback can play a significant role in motivating students to be independent, active drivers for their own academic learning (Selvaraj et al., 2021), peer assessment stimulates learners' self-confidence, active engagement, participation, and autonomy (López et al., 2022; Prosenjak & Lučev, 2020; Yundayani et al., 2024). It also fosters a supportive classroom environment and social cohesion (Maiz Arévalo, 2008). Peer feedback can increase understanding of assessment criteria and quality standards (McGarrigle, 2013). Consistent longitudinal practice and training have been associated with students' perception of improved peer assessment skills and quality (Gudiño et al., 2024). According to Panadero et al. (2023), peer assessment is influenced by some internal factors like self-efficacy, motivation, level of comfort, or fairness perception, along with other external factors like social relationships, perceived psychological safety, or trust in others as evaluators. Higher comfort has been reported when peer feedback is given anonymously (Lozano Zumba et al., 2025; Su, 2023). Anonymous peer assessment appears to enhance students' awareness of its learning value, promoting more critical feedback and improving their performance, particularly in higher education settings (Panadero & Alqassab, 2019).

Using technology like mobile apps to support peer feedback in the EFL classroom offers clear advantages (Charoensuk & Wanchid, 2025; Gokgoz-Kurt, 2023; Yundayani et al., 2024). Some studies have shown that integrating technological tools can reduce negative social pressures, such as peer pressure and fear of disapproval, by enabling anonymous feedback, resulting in more positive student perceptions and potentially more valid assessment outcomes (Wu & Miller, 2020). Yet, these investigations have also reported concerns about teacher control, classroom dynamics, small device screens, and limited feedback options, suggesting that while technology enhances objectivity and engagement, it also introduces some constraints that need to be considered.

Systematic reviews have confirmed that peer feedback, when structured and supported, can boost learning outcomes and student engagement in EFL settings (Prosenjak & Lučev, 2020). The benefits of peer feedback for oral skill development in the EFL classroom have been empirically demonstrated (Au & Bardakçı, 2020; Chekol, 2020; Le et al., 2025). In EFL higher education environments, this type of feedback can contribute to enhancing students' pronunciation, fluency, and vocabulary (Le et al., 2025). Peer feedback contributes to developing oral presentation skills (Prosenjak & Lučev, 2020; Suharni et al., 2022; Wanchid & Charoensuk, 2024; Yüce & Curle, 2025). A study carried out by Gokgoz-Kurt (2023) explored peer feedback given by EFL students on their classmates' oral presentation performances through an online platform. The results showed that most feedback was positive rather than negative, focusing on content, spoken performance, and presentation skills. While students raised concerns about friendship bias, anonymity, and tone of feedback, they generally considered mobile-assisted peer feedback as a valuable and effective tool.

Evaluations from both peers and teachers greatly impact oral presentation development in EFL (Charoensuk & Wanchid, 2025), and each type of feedback has a different effect on learners' presentation performance (Wanchid & Charoensuk, 2024). Research indicates that combining peer and teacher assessment can provide a more holistic and reliable evaluation of student performance, as well as encouraging students to value feedback from multiple sources, not just the teacher (Maiz Arévalo, 2008; Salehi & Gholampour, 2022). This can motivate students to participate more fully in the learning process (Nejad & Mahfoodh, 2019), especially when peer grades contribute to final grades (Maiz Arévalo, 2008; McGarrigle, 2013).

Wanchid and Charoensuk (2024) found that both peer and teacher feedback significantly improved EFL students' oral presentation performance, with students expressing positive perceptions of peer feedback despite noting some challenges. Other scholars have demonstrated that providing double assessment not only reinforces the teacher's evaluation but also increases student motivation and involvement (Liu & Aryadoust, 2024; Maiz Arévalo, 2008). Another study conducted by Nejad and Mahfood (2019) on the effectiveness of self-, peer-, and teacher assessments in EFL students' oral presentations concluded that peer and teacher assessments were consistent. Teacher feedback was regarded as more accurate and useful by the students, and an examination of the mean scores proved that teachers employed scoring criteria more strictly (Nejad & Mahfood, 2019).

However, peer assessment in oral presentations may also lead to presentation anxiety and difficulties in feedback provision and scoring (Widodo & Chakim, 2023). Other challenges include potential bias, varying levels of peer expertise, and cultural reluctance to critique peers (Wanchid & Charoensuk, 2024). Providing students with training and clear rubrics is fundamental to ensure constructive and reliable peer feedback (Charoensuk & Wanchid, 2025; McGarrigle, 2013; Okumu et al., 2024; Wanchid & Charoensuk, 2024). Teachers should be trained in peer assessment as well, to encourage them to use it effectively (Le et al., 2025). In addition, achievement levels should be taken into account since advanced students may need more elaborate feedback, whereas lower levels may require more targeted language support (Charoensuk & Wanchid, 2025).

**RESEARCH METHOD**

This research was derived from a pedagogical intervention in the Bachelor's Degree in Tourism at CETT-Barcelona School of Tourism, Hospitality and Gastronomy, affiliated to the University of Barcelona, in Spain. The participants had been instructed to undertake a research project in pairs in the second semester of their 90-hour EFL course. The project guidelines were published on their virtual campus, and a class blog was also used as a teaching-learning-assessment platform: Welcome to CETT's blog (http://englishcett.blogspot.com). Detailed instructions were posted on the blog, coupled with occasional peer and teacher feedback, with students' consent.

Students had worked on the initial stages of their project and were required to present their progress orally. In line with best practices recommended by other authors (Salehi & Gholampour, 2022), detailed criterion-referenced rubrics were clearly explained to students from the beginning of the academic period, and in prior oral presentations, they were trained to use them and comment on their peers' presentations, complemented by the teacher's feedback. However, this was the first time that their scores were taken into account for grading purposes.

The assessment criteria were:

1. Language use (grammar/vocabulary accuracy, range & adequacy).
2. Task achievement (efficient achievement of communicative purpose, content relevance & extent).
3. Fluency and pronunciation (speech flow, intonation, stress & rhythm).
4. Communicative resources (use of linguistic and non-linguistic resources, supporting material, attitude).

For this oral presentation, students were asked to scan a QR code and fill in a Google Form anonymously for each individual presenter (except for their own team) using their laptops, tablets or cell phones. They had to rate each criterion on the following scale: poor, fair, satisfactory, very good, and excellent. Students were also asked to write a positive aspect and an improvement suggestion in relation to every performance. The teacher explained that their ratings would be turned into a class grade for each student, which would be combined with the teacher's rating and result in each student's final grade for this task. For this purpose, the teacher exported the anonymous Google Form responses and compiled them in Microsoft Excel. Peer ratings for each criterion were then converted into numerical values on a 10-point Likert scale as follows: poor=2, fair=4, satisfactory=6, very good=8, and excellent=10. Mean peer rates were calculated. The teacher, who had previously rated each performance by herself on a 10-point Likert scale, then calculated the mean value resulting from her own rating and the mean peer rating. Finally, the teacher informed each student of the final grade they had obtained and also shared with them the feedback provided by their classmates.

This descriptive study has a mixed-methods approach. Google Form responses sent by 50 students were collected throughout three academic years, namely 2020-2021, 2023-2024, and 2024-2025. If all the students had commented on the presentations of all their classmates, excluding those of their own team, the total number of responses would have been 752. Eventually, 341 forms were collected, which represents a margin of error of 3.93% and a confidence level of 95% with simple random sampling and assuming maximum variability ($p$ = 0.5). The peer feedback forms included both Likert-scale ratings on a 0–10 scale and open-ended qualitative comments for each of the four assessment criteria: language use, task achievement, fluency and pronunciation, and communicative resources. Each student was assessed by multiple peers and the teacher, and their final grade was calculated as the average

of all ratings received, as mentioned above. Table 1 shows the characteristics of the participants and the data set.

**Table 1** Participants and data set

| |
|---|
| Total participants N = 50 undergraduate students |
| 2020-21 n = 16 (S1-S16)<br>2023-24 n = 20 (S17-S36)<br>2024-25 n = 14 (S37-S50) |
| Total responses N = 752 |
| 2020-21 n = 224<br>2023-24 n = 360<br>2024-25 n = 168 |
| Responses collected N = 341 |
| 2020-21 n = 97<br>2023-24 n = 132<br>2024-25 n = 112 |

Salehi and Gholampour (2022) state that to guarantee reliability in peer assessment, internal consistency among peer raters for each presentation should be calculated using Cronbach's alpha, considering that scores ≥ 0.70 are acceptable. According to these authors, the validity of peer ratings in relation to those of the instructor can be ensured by looking at the differences in standard deviations of both peers' and the teacher's mean scores. Peer ratings are valid if their mean scores are within one standard deviation of the teacher's mean score. Additionally, correlational analyses should be conducted between both scores to identify significant positive correlations, taking teacher assessments as the gold standard for criterion-related validity (Salehi & Gholampour, 2022).

To explore the alignment between peer and teacher scores, two main statistical analyses were performed following Salehi and Gholampour (2022): (a) inter-rater reliability was assessed using Cronbach's alpha and standard deviation of score differences; and (b) Pearson correlation coefficients were calculated to measure the strength of the relationship between peer and teacher grades for each academic year.

To conduct the descriptive and inferential statistical analyses, raw peer scores were exported from Google Forms and compiled in Microsoft Excel. For each student, average peer ratings across all categories were computed, as well as the corresponding teacher scores. These data were then processed in GNU PSPP to calculate means, standard deviations, Pearson correlation, and Cronbach's alpha. On the other hand, a qualitative analysis of the comments in regard to strengths and weaknesses was performed by means of Julius.ai, categorizing the aspects mentioned by all the students in their open comments and counting their occurrences.

## RESULTS AND DISCUSSION

### *Quantitative Analyses*

This section reports the results of the statistical analysis comparing peer and teacher assessments across three academic years: 2020–2021, 2023–2024, and 2024–2025. Across the three cohorts, peer-assigned grades for each assessment criterion remained within a relatively narrow range. Figure 1 illustrates this.

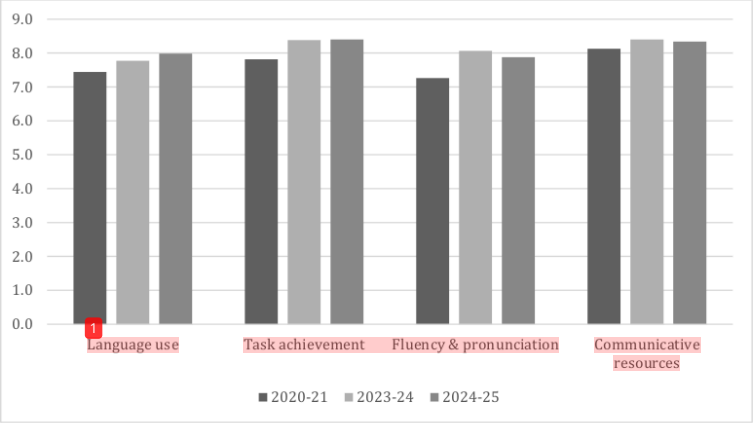


**Figure 1** Peer ratings across years

In 2020–21, the highest mean score is for communicative resources (8.1); in 2023-24, for both communicative resources and task achievement (8.4); and in 2024-25, for task achievement, as shown in Table 2. In 2020–21, the mean peer grade (M) was 7.7 with a standard deviation (SD) of 1.9; in 2023–24, the mean grade rose (M = 8.2, SD = 1.5) and remained similar in 2024–25 (M = 8.1, SD = 1.6). The distribution of peer grades for each year shows a skew towards the upper end of the scale, especially in 2023–24, where grade inflation was most pronounced.

**Table 2** Peer mean values and standard deviations

| | 2020-21 | | 2023-24 | | 2024-25 | |
|---|---|---|---|---|---|---|
| | **M** | **SD** | **M** | **SD** | **M** | **SD** |
| Language use | 7,4 | 1,7 | 7,8 | 1,6 | 8,0 | 1,5 |
| Task achievement | 7,8 | 1,8 | 8,4 | 1,4 | 8,4 | 1,5 |
| Fluency & pronunciation | 7,3 | 1,9 | 8,1 | 1,4 | 7,9 | 1,9 |
| Communicative resources | 8,1 | 2,0 | 8,4 | 1,4 | 8,3 | 1,6 |
| M peer grade | 7,7 | 1,9 | 8,2 | 1,5 | 8,1 | 1,6 |

Although these scores suggest consistency, closer examination revealed differences in alignment with teacher grades. To measure alignment, scores were organized into a student-per-rater matrix for each cohort, and individual discrepancies between the average peer score and the teacher's score for each student were calculated. Figure 2 illustrates that in 2020–21 the teacher awarded slightly higher grades (M = 8.1) than peers (M = 7.7), producing a positive mean difference of +0.4. Therefore, the teacher tended to evaluate student performance more favorably than peers across the sample. Even though the standard deviation of differences was low (SD = -0.5) and reflected strong alignment, peer grades showed slightly greater variability (SD = 1.2) relative to the teacher's grades (SD = 0.8). This suggests that peers applied a wider range of scoring, potentially due to heterogeneous judgment criteria.

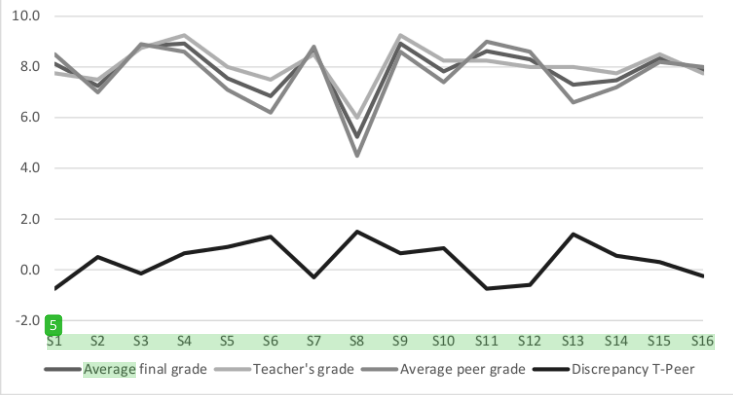


**Figure 2** Academic year 2020-21

At the individual level, discrepancies varied considerably, ranging from strong overestimation by peers (e.g., S11 and S1, both at –0.8) to substantial underestimation relative to the teacher (e.g., S8 at 1.5 and S13 at 1.4). These larger discrepancies may indicate cases in which peer assessors either misinterpreted the assessment criteria or were influenced by extraneous factors such as interpersonal dynamics, perceived effort rather than quality, or limited assessment literacy. Students with lower performance levels (e.g., S8, S6, S13) tended to receive markedly harsher peer grades than teacher grades, which may mean that peers penalized weaker work more severely owing to less calibrated expectations. Conversely, for higher-performing students (e.g., S3, S7, S11), peer overestimation was slightly more common. This may be caused by a halo effect, wherein generally strong performers receive marginally inflated evaluations from classmates. Despite this, the magnitude of overestimation remained relatively small for most high-achieving cases. Table 3 shows the matrix for 2020-21.

**Table 3** Individual values for 2020-21

| Student | Average final grade | Teacher's grade | Average peer grade | Discrepancy T-Peer |
|---|---|---|---|---|
| S1 | 8,1 | 7,8 | 8,5 | -0,8 |
| S2 | 7,3 | 7,5 | 7,0 | 0,5 |
| S3 | 8,8 | 8,8 | 8,9 | -0,2 |
| S4 | 8,9 | 9,3 | 8,6 | 0,7 |
| S5 | 7,6 | 8,0 | 7,1 | 0,9 |
| S6 | 6,9 | 7,5 | 6,2 | 1,3 |
| S7 | 8,7 | 8,5 | 8,8 | -0,3 |
| S8 | 5,3 | 6,0 | 4,5 | 1,5 |
| S9 | 8,9 | 9,3 | 8,6 | 0,7 |
| S10 | 7,8 | 8,3 | 7,4 | 0,9 |
| S11 | 8,6 | 8,3 | 9,0 | -0,8 |
| S12 | 8,3 | 8,0 | 8,6 | -0,6 |
| S13 | 7,3 | 8,0 | 6,6 | 1,4 |
| S14 | 7,5 | 7,8 | 7,2 | 0,6 |
| S15 | 8,4 | 8,5 | 8,2 | 0,3 |
| S16 | 7,9 | 7,8 | 8,0 | -0,3 |
| M | 7,9 | 8,1 | 7,7 | 0,4 |
| SD | 1,0 | 0,8 | 1,2 | -0,5 |

In contrast, the 2023–24 cohort displayed a markedly different pattern, which can be observed in Figure 3, with peers consistently assigning higher grades than the teacher. The mean peer grade (M = 8.2) exceeded the mean teacher grade (M = 6.9) by a substantial margin, resulting in a pronounced negative mean discrepancy of –1.3. This differed from the 2020-21 cohort, in which teacher grades were generally higher. The shift suggests that the reliability and directionality of peer judgments may be highly context-dependent, influenced by cohort characteristics, task type, group dynamics, or the degree of scaffolding provided for the assessment process.

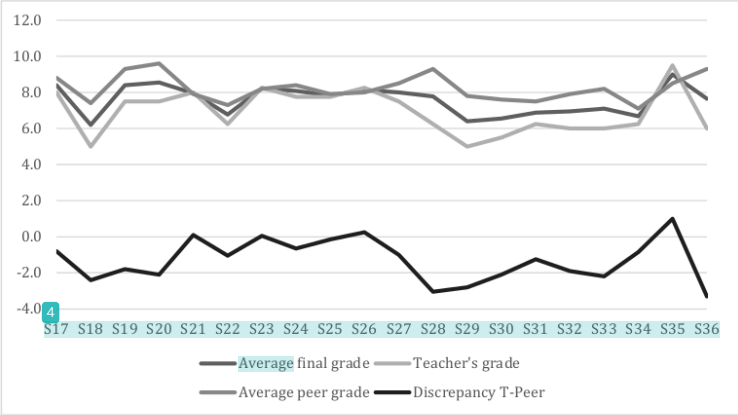


**Figure 3** Academic year 2023-24

As portrayed in Table 4, the magnitude of discrepancies in this second set was considerably larger, with several students receiving peer grades that exceeded teacher grades by more than two full points (e.g., S18 at –2.4, S20 at –2.1, S28 at –3.1, S29 at –2.8, S36 at –3.3). Such deviations imply a systemic upward bias in peer assessment that may reflect leniency, reciprocal altruism, or reluctance to critically evaluate classmates' work. In some cases, this inflation reached levels that would meaningfully distort summative judgments if peer assessments were used without moderation. The standard deviation of teacher grades (SD = 1.2) was higher than that of peer grades (SD = 0.7), indicating that the teacher discriminated more strongly between levels of student performance. The reduced variability may have stemmed from uncertainty regarding quality thresholds, limited domain expertise, or a desire to avoid interpersonal tension, all of which tended to push peer scores toward homogeneity. Individual cases further illuminate these dynamics. For instance, students with lower teacher-assessed performance (e.g., S18, S29, S30, S31) received disproportionately high peer evaluations, suggesting that peers may have either overvalued effort or lacked the knowledge to detect deficiencies. On the contrary, S35 was the only notable case of teacher grades exceeding peer grades (1.0). This may denote that peers undervalued strong performance, perhaps because they misinterpreted some aspects that the teacher recognized as strengths.

**Table 4** Individual values for 2023-24

| Student | Average final grade | Teacher's grade | Average peer grade | Discrepancy T-Peer |
|---|---|---|---|---|
| S17 | 8,4 | 8,0 | 8,8 | -0,8 |
| S18 | 6,2 | 5,0 | 7,4 | -2,4 |
| S19 | 8,4 | 7,5 | 9,3 | -1,8 |
| S20 | 8,6 | 7,5 | 9,6 | -2,1 |
| S21 | 8,0 | 8,0 | 7,9 | 0,1 |
| S22 | 6,8 | 6,3 | 7,3 | -1,1 |
| S23 | 8,2 | 8,3 | 8,2 | 0,1 |
| S24 | 8,1 | 7,8 | 8,4 | -0,7 |
| S25 | 7,8 | 7,8 | 7,9 | -0,2 |
| S26 | 8,1 | 8,3 | 8,0 | 0,3 |
| S27 | 8,0 | 7,5 | 8,5 | -1,0 |
| S28 | 7,8 | 6,3 | 9,3 | -3,1 |
| S29 | 6,4 | 5,0 | 7,8 | -2,8 |
| S30 | 6,6 | 5,5 | 7,6 | -2,1 |
| S31 | 6,9 | 6,3 | 7,5 | -1,3 |
| S32 | 7,0 | 6,0 | 7,9 | -1,9 |
| S33 | 7,1 | 6,0 | 8,2 | -2,2 |
| S34 | 6,7 | 6,3 | 7,1 | -0,9 |
| S35 | 9,0 | 9,5 | 8,5 | 1,0 |
| S36 | 7,7 | 6,0 | 9,3 | -3,3 |
| M | 7,6 | 6,9 | 8,2 | -1,3 |
| SD | 0,8 | 1,2 | 0,7 | 1,2 |

The 2024–25 cohort presented a markedly higher degree of alignment between teacher and peer assessments, shown in Figure 4. Both the mean teacher grade (M = 8.1) and the mean peer grade (M = 8.2) were closely matched, resulting in a negligible mean discrepancy of –0.1. The students and the teacher seemed to operate with a broadly shared understanding of performance standards. Such consistency is noteworthy, as it indicates conditions conducive to reliable peer assessment, whether through clear rubrics, prior experience with evaluative tasks, or cohesive group norms. The standard deviations for teacher (SD = 0.9) and peer grades (SD = 0.8) were also closely aligned, indicating similar levels of score dispersion. This suggests that peers, unlike in the 2023-24 cohort, did not exhibit pronounced leniency. Instead, their evaluations differentiated between varying levels of performance in a manner comparable to the teacher assessments. Although some discrepancies remained, they generally fell within a modest range (–0.8 to 0.8), indicating limited divergence in judgment.

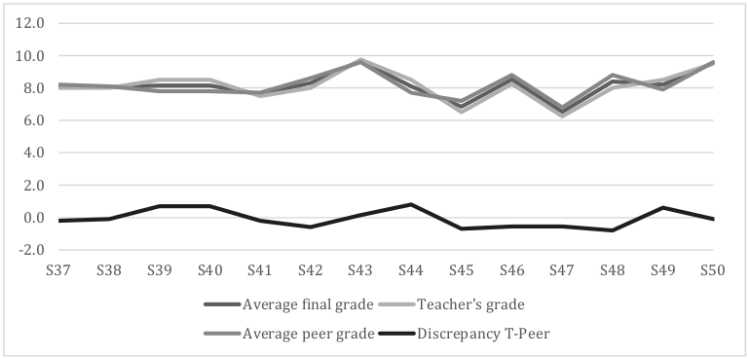


**Figure 4** Academic year 2024-25

A closer examination of the individual cases gathered in Table 5 revealed both directions of discrepancy, but at moderate magnitudes. Students such as S39, S40, and S44 displayed positive discrepancies (0.7–0.8), so the teacher's assessments were somewhat more generous than peer evaluations. These cases may mean that the students applied more conservative interpretations of the criteria or were more attentive to minor aspects overlooked by the teacher. Conversely, students such as S48 (–0.8) and S42, S46, S47 (all –0.6) received slightly higher peer than teacher grades, though again the differences were small. These minor instances of peer inflation do not appear systematic but may reflect typical inter-rater variability, found even among trained evaluators. Importantly, several cases of near-perfect alignment can be identified, such as S37, S38, S41, and S50, where discrepancies were minimal (–0.2 to –0.1). Such consistency suggests that peers in this cohort were able to apply evaluative criteria with a degree of precision comparable to that of the teacher. The extraordinarily high performance of S43 and S50, with scores above 9.5 from both the teacher and peers, further pointed to shared recognition of exceptional quality performance.

**Table 5** Individual values for 2024-25

| Student | Average final grade | Teacher's grade | Average peer grade | Discrepancy T-Peer |
|---|---|---|---|---|
| S37 | 8,1 | 8,0 | 8,2 | -0,2 |
| S38 | 8,1 | 8,0 | 8,1 | -0,1 |
| S39 | 8,2 | 8,5 | 7,8 | 0,7 |
| S40 | 8,2 | 8,5 | 7,8 | 0,7 |
| S41 | 7,6 | 7,5 | 7,7 | -0,2 |
| S42 | 8,3 | 8,0 | 8,6 | -0,6 |
| S43 | 9,7 | 9,8 | 9,6 | 0,2 |
| S44 | 8,1 | 8,5 | 7,7 | 0,8 |
| S45 | 6,9 | 6,5 | 7,2 | -0,7 |
| S46 | 8,5 | 8,3 | 8,8 | -0,6 |
| S47 | 6,5 | 6,3 | 6,8 | -0,6 |
| S48 | 8,4 | 8,0 | 8,8 | -0,8 |
| S49 | 8,2 | 8,5 | 7,9 | 0,6 |
| S50 | 9,6 | 9,5 | 9,6 | -0,1 |
| M | 8,2 | 8,1 | 8,2 | -0,1 |
| SD | 0,8 | 0,9 | 0,8 | 0,6 |

To examine in more depth the alignment between peer and teacher evaluations, Pearson correlation coefficients were calculated for each academic year having set an alpha level of .05. Results showed a strong positive correlation in both 2020–21 ($r(14) = .84$, $p < .001$) and 2024–25 ($r(12) = .80$, $p < .001$), while a moderate correlation was observed in 2023–24 ($r(18) = .56$, $p < .001$). Some key considerations for validity must be made at this point. Correlations were interpreted considering that each student in the cohort received a minimum of 4 feedback responses. Peer scores were averaged per student across multiple ratings before correlating with the teacher's single score. Moreover, all scores were based on the same rubric with the abovementioned consistent assessment criteria, i.e., language use, task achievement, fluency and pronunciation, and communicative resources.

These findings indicate that peer assessments were generally aligned with teacher evaluations, as observed in Nejad and Mahfood's (2019) study. However, the discrepancies in the 2023–24 cohort reflected a lower level of consistency and possibly indicated a lack of scoring accuracy. It is likely that, in this specific case, the teacher applied stricter scoring criteria, as also reported by Nejad and Mahfood (2019).

To evaluate internal consistency, Cronbach's alpha was calculated for the peer scores in each year. The results indicate high inter-rater reliability in 2020–21 ($\alpha = 0.89$) and 2024–25 ($\alpha = 0.87$), exceeding the threshold of 0.70 for high reliability designated by Salehi and

Gholampour (2022). In contrast, the lower reliability in the 2023–24 cohort ($\alpha = 0.72$) indicates moderate internal consistency, pointing to greater scoring variability in peer evaluations. This result is consistent with additional analyses showing that several students in 2023–24 received peer scores even 3 points higher than their teacher-assigned scores, suggesting possible overrating or rubric misalignment.

These results offer several important insights. First, they suggest that with appropriate scaffolding and familiarity with assessment criteria, peer assessment can be both reliable and valid, as stated by other authors (Lozano Zumba et al., 2025; Maiz Arévalo, 2008), closely mirroring teacher evaluations. The strong correlation and high inter-rater reliability in 2020–21 and 2024–25 confirm that students in those years were able to apply the grading rubric with reasonable consistency. This indicates that with appropriate guidance, peer ratings can closely reflect teacher judgments, enhancing the validity of peer-assigned grades in summative oral assessment. Although the inclusion of peer data in grade aggregation can exert an upward pull under conditions of indulgent peer scoring, the integration of multiple assessment sources can also balance high-stakes evaluations. As already mentioned, the overestimation in 2023–24 may have been due to the cultural reluctance to criticize other students' performance reported by Wanchid and Charoensuk (2024). It may also reflect leniency, lack of assessment literacy, or insufficient rater training. These weaker results indicate that instructor moderation and norms concerning academic honesty are required when peer grades are used for summative purposes. Moreover, they underscore the need for calibration activities, enhanced training, and clearer rubrics, in line with previous research (Charoensuk & Wanchid, 2025; McGarrigle, 2013; Okumu et al., 2024; Wanchid & Charoensuk, 2024).

***Qualitative Analysis***

The analysis of open-ended peer feedback provides a clear overview of the key themes identified in the comments, which were counted and categorized. A total of 200 mentions of strengths were detected. Fluency emerged as the most frequently perceived strength, accounting for 27.5% of all mentions, followed by vocabulary at 17.5%. Other notable strengths included pronunciation (9.5%), confidence/attitude (9.0%), and communication/presentation skills (8.5%). Lower percentages were observed for knowledge/understanding (6.0%) and visual aids/design (5.5%), while grammar and accuracy were rarely mentioned, representing only 2.0% and 1.5%, respectively. Below is a sample of some literal peer feedback comments:

*"She has a really good fluency."*
*"She used a lot of vocabulary."*
*"Good pronunciation."*

On the other hand, a total of 155 mentions of weaknesses were identified. The most frequently cited weakness was fluency, representing 16.1% of all mentions, closely followed by pronunciation at 15.5%. Other significant areas of concern included vocabulary (12.3%), reading/not speaking directly (10.3%), and nervousness/lack of confidence (10.3%), indicating that delivery-related issues were among the most common challenges. Lower percentages were noted for grammar and non-verbal communication, each accounting for 6.5%, while lack of eye contact was mentioned in 5.2% of cases. Minor weaknesses included monotonous presentation, presentation design, and lack of information/details, each at 3.2%. The following literal comments illustrate some of the main weaknesses perceived by the students:

*"She needs to improve a little the fluency."*
*"Maybe the pronunciation could have been better."*
*"She needs more self confidence in herself, she has more level than she thinks."*
*"He gets nervous and that plays against him... Keep calm, you're doing good :)"*

Globally, the results align with the findings described by Gokgoz-Kurt (2023) in terms of focus on oral and presentation skills. In our study, fluency, vocabulary, and pronunciation were the most frequently mentioned strengths. Interestingly, fluency and pronunciation also appeared as the most common weaknesses, indicating that these dimensions may be particularly salient for students. Other frequently cited areas for improvement included aspects of presentation delivery, such as nervousness or lack of confidence. These qualitative trends remained consistent across academic years, reinforcing the credibility of the feedback patterns outlined by the quantitative analysis. The bar charts in Figure 5 visually represent the counts of each category for both strengths and weaknesses to understand the distribution of comments across different themes.

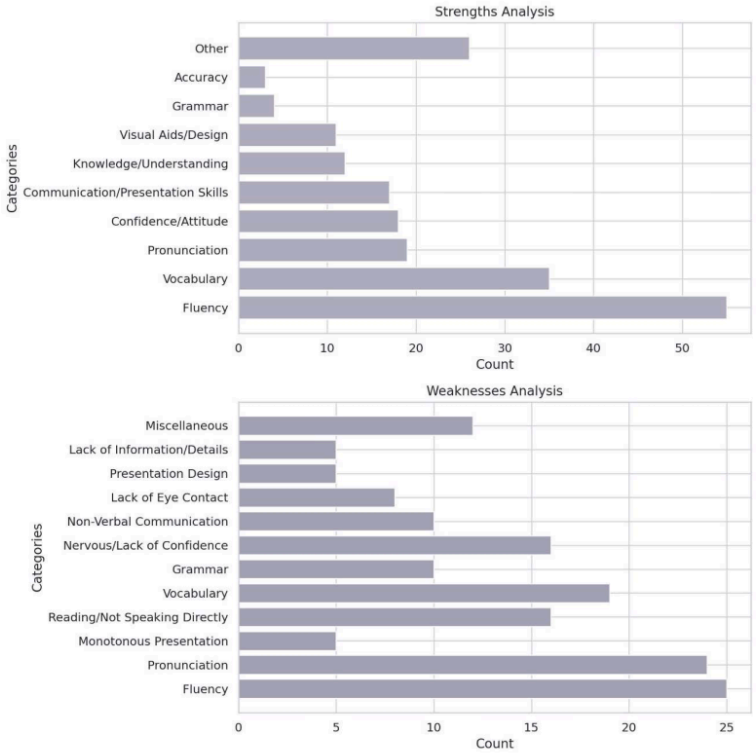


**Figure 5** Categorization of strengths and weaknesses

To sum up, it can be observed that, as concluded by Gokgoz-Kurt (2023), learners also made a higher number of positive comments. Nevertheless, the wider range of negative aspects

pinpointed in our study denotes the students' ability to detect their peers' deficits critically and supports most of our quantitative results.

From a socio-constructivist perspective, the combined quantitative and qualitative findings of this study denote that the co-construction of knowledge in assessment does not take place only among students. Grounded on Vygotsky's principles regarding the importance of learning through social interaction (Iglesias, 2013; Janesarvatan & Asoodar, 2024; McGarrigle, 2013), our results show that assessment can be more than a mechanism for grading. It can become a shared learning activity in which teachers and students build common standards together. This extends socio-constructivist conceptions by highlighting that social learning includes both horizontal interactions among peers and vertical interactions with instructors, particularly if they have the opportunity to compare, negotiate, and reflect on the differences between peer and teacher feedback to enhance their understanding and development of stronger evaluative skills.

This notion also contemplates a broader shift in how collaborative learning may evolve in the near future. As digitalization and artificial intelligence (AI) tools increasingly take part in educational practice and feedback processes, they can add a new voice to the dialogue that shapes students' learning. From a Vygotskian point of view, AI can be seen as another mediating tool within the learner's zone of proximal development that supports learners by modelling good performance, prompting reflection, or helping them identify areas for improvement interactively. Bringing peers, teachers, and AI together in the feedback process has the potential to create a richer and more dynamic learning environment in which each contributor helps students build more accurate judgment and a deeper understanding. This suggests that socio-constructivist theory may need to expand to account for these new forms of mediated interaction, where the co-construction of knowledge is shared across a wider network of participants, human or technological.

## CONCLUSION

The combined use of descriptive and inferential statistics, coupled with content analysis of peer feedback comments, has made it possible to assess the trustworthiness of mobile-assisted peer evaluations and to better understand the dynamics of collaborative assessment in EFL contexts. The results of this study indicate that peer feedback consistently identified fluency, vocabulary, and pronunciation as both strengths and areas for improvement. Peer scores aligned reasonably well with teacher scores in 2020–21 and 2024–25, suggesting moderate to strong validity. Yet, 2023–24 showed inflated peer ratings, likely due to overgenerous evaluation or insufficient peer assessment training. This research contributes to the underexplored area of integrating peer and teacher assessment in EFL oral presentations. It provides empirical evidence on peer-teacher rating alignment and offers insight into peer comment trends across multiple cohorts. Therefore, this study supports the integration of peer assessment in higher education and offers guidelines for ensuring its effective implementation.

The practical implications entail that combined peer and teacher assessment can be valid for oral presentation grading, especially with adequate training and clear rubrics. Peer feedback enhances student engagement and self-awareness when structured effectively, so instructors and educational institutions should incorporate them in their programs actively. Nevertheless, fairness and consistency are relevant concerns that also need to be addressed.

This study aligns with the socio-constructivist theoretical conception that the co-construction of knowledge in assessment contexts extends beyond peer-to-peer interaction. The refining and calibration of evaluative judgment through both peers and teachers suggests

that assessment functions as a socially mediated learning activity wherein feedback processes serve as opportunities for cognitive scaffolding, mutual regulation, and the development of shared standards of quality. Looking ahead, another implication of this study is that socio-constructivist frameworks must evolve to account for triangulated feedback among peers, teachers, and AI systems to reconfigure the social dynamics of assessment, where human and technological contributions collectively shape learners' evaluative reasoning and academic development.

The small sample in this study does not allow for significant representativeness and extrapolation. Results are based on a single educational context, so generalizability may be limited. Another limitation is related to assessment validity and reliability. According to Salehi and Gholampour (2022), by combining quantitative methods with structured rubrics and ongoing training, educators can ensure peer feedback in oral presentations is both reliable and valid. At least four peers should rate each presentation to improve reliability, which can be tracked longitudinally to identify trends (e.g., improved consistency over time with rubric familiarity). Moreover, running independent t-tests and comparing scores from high- vs. low-proficiency peers can contribute to checking for bias, while ANOVA tests can be used to assess whether presentation topics affect scoring patterns.

Future research could examine students' perceptions of both the process and outcomes of peer feedback and assessment in oral presentations. Additionally, it would be valuable to study whether integrating peer and teacher assessments enhances students' sense of fairness regarding the grading of oral presentations. Another important aspect to investigate is the impact of combining peer and teacher assessments on the development of students' oral presentation skills in EFL higher education over time. Finally, exploring how technology can be leveraged to streamline the peer feedback process could provide significant insights for improving educational practices.

**REFERENCES**

# Integrating Peer and Teacher Feedback in the Assessment of English as a Foreign Language Oral Presentations in Higher Education

ORIGINALITY REPORT

| 4% | 3% | 2% | 1% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | files.eric.ed.gov<br>Internet Source | 1% |
| 2 | www.researchgate.net<br>Internet Source | 1% |
| 3 | www.cett.es<br>Internet Source | 1% |
| 4 | c.coek.info<br>Internet Source | <1% |
| 5 | peer.asee.org<br>Internet Source | <1% |
| 6 | reliefweb.int<br>Internet Source | <1% |
| 7 | "Academic Tourism", Springer Science and Business Media LLC, 2021<br>Publication | <1% |
| 8 | longislandartsalliance.org<br>Internet Source | <1% |
| 9 | Winnie Cheng, Martin Warren. "Peer and Teacher Assessment of the Oral and Written Tasks of a Group Project", Assessment & Evaluation in Higher Education, 1999<br>Publication | <1% |
| 10 | www.mdpi.com<br>Internet Source | |

<1%

| | | | |
|---|---|---|---|
| Exclude quotes | Off | Exclude matches | Off |
| Exclude bibliography | Off | | |