

Academic Report

Integrating Peer Assessment into Speaking and Listening Courses: A Structured Classroom Approach

Simon G. WILKINS and Peter J. COLLINS
(wilkins@tokai.ac.jp; pcollins@tokai.ac.jp)
Tokai University

Abstract

This academic report describes a structured, classroom-tested approach to integrating peer assessment into tertiary-level English communication courses emphasizing listening and speaking skills. Drawing on formative assessment theory and learner autonomy frameworks, the authors outline a four-stage process for implementing a peer-assessment framework that is both pedagogically grounded and user-friendly for both teachers and students. The framework's approach incorporates learner-generated criteria, authentic benchmarking, and technology-enhanced feedback. Insights from classroom applications and reflections on the authors' own experiences developing and implementing this framework suggest that peer assessment may not only improve communicative performance but also foster metacognitive awareness and support the development of assessment-capable learners. This academic report integrates existing research into a practical classroom framework.

Keywords: peer assessment, CALL, Assessment for Learning, speaking, lifelong learning

1. Introduction

The purpose of this paper is to offer a literature-informed approach that draws on previous findings on peer assessment in speaking and listening courses. It presents a practical, adaptable framework with which readers can imagine ways to implement peer assessment in their own teaching contexts. Examples from and reflections on the authors' own experiences using peer assessment are included that will help readers predict and overcome potential obstacles to making peer assessment a meaningful part of their own courses.

1.1 Context and Course Setting

This framework was developed in elective English communication courses at a private university in Japan. Courses featured in this study run for 14 weeks across one semester and include 25–30 students per class representing a range of non-English majors. Student English proficiency levels range from CEFR A1–B2, and the courses emphasize listening and speaking. The peer-assessment framework has been implemented across multiple cohorts over several academic semesters, allowing for iterative refinement based on classroom experience. All the courses use a teacher-administered Moodle LMS with an H5P Interactive Video plugin enabled for student use. Students complete recordings on personal smartphones, tablets, or laptops, reflecting the university's bring-your-own-device environment. This context influences both the technological feasibility and the

pedagogical design of the framework described in this report.

1.2 Background and Rationale

Peer assessment as a framework of listening–speaking activities has gained increasing attention in the past decade in language learning education as a formative strategy that promotes learner autonomy, metacognitive awareness, and communicative competence. Traditionally, speaking assessment focused on formal proficiency testing (Luoma, 2004; McNamara, 1996), but recent pedagogical shifts have emphasized learning-oriented assessment where the primary goal is to enhance learning through active learner involvement in the assessment process (Bennett, 2024; Carless & Winstone, 2020).

Self- and peer assessment of speaking, where learners evaluate their own and others' oral performance, has emerged as a powerful tool for formative learning (Saito, 2008; Topping, 2018). These practices encourage learners to engage critically with performance criteria, reflect on their communicative strengths and weaknesses, and internalize standards of effective communication (Boud & Soler, 2016; Carless, 2015).

Peer assessment of speaking presents a range of challenges that may make teachers hesitant to adopt it, particularly in contexts constrained by curriculum design, class size, time, and/or assessment logistics (Frey, Fisher, & Hattie, 2018). Without adequate preparation, peer assessment can appear unreliable, overly demanding for learners, and difficult to manage fairly.

Teachers and learners may also perceive peer assessment as less authoritative than teacher feedback, especially when institutional accountability measures emphasize standardized evaluation (Double, McGrane, & Hopfenbeck, 2020). One concern is the inconsistency of learner judgments. Learners being asked to assess their peers need shared standards of quality and clear language goals. Without calibrated exemplars and co-constructed criteria, their evaluations tend to be idiosyncratic and to drift over time (Liu & Carless, 2006; Nicol & Macfarlane-Dick, 2006; Sadler, 1989). Inexperienced assessors often apply implicit standards or rely on personal preferences rather than shared benchmarks of quality (Gielen, Dochy, & Onghena, 2011; Saito, 2008). This variability may be further compounded by cultural conceptions of politeness and directness in feedback (Hyland & Hyland, 2006), leading to distortions in how learners phrase or interpret evaluative comments.

Another consideration is the cognitive load of dual processing. Listening while simultaneously evaluating peers' speech imposes substantial cognitive demands that can reduce attention and accuracy (Skehan, 1998; Sweller, 1994).

Finally, certain classroom dynamics can complicate the process. When learners are both giving and receiving feedback, role conflict, reciprocity pressures, and reluctance to offer critical comments can undermine validity and fairness (Falchikov & Goldfinch, 2000; Ross, 1998; Topping, 2009). In an age when AI tools deliver instant automated evaluations, peer feedback may feel slow and uncertain, reducing motivation to engage meaningfully with it.

Far from invalidating the use of peer assessment, these challenges highlight the need for structured scaffolds such as explicit criteria, anonymization where possible, and staged practice with exemplars to ensure reliability and fairness. Despite obstacles, research indicates that limitations to peer assessment can be effectively mitigated through structured design and targeted support. Clear rubrics, co-constructed criteria, and

exposure to benchmark exemplars provide shared standards of quality and explicit language goals (Nicol & Macfarlane-Dick, 2006; Sadler, 1989). Empirical studies confirm that, when these supports are in place, peer assessment can yield reliable and valid results: Saito's (2008) meta-analysis found a moderate correlation between peer and teacher ratings, and Bachman and Palmer (1989) showed that peer and self-assessments can validly measure communicative competence when aligned with well-defined linguistic traits. Scaffolding tasks into manageable steps helps focus learner attention on specific features of speaking performance (Skehan, 1998; Sweller, 1994), while assessor training, calibration, and multiple practice rounds under anonymized conditions enhance reliability and fairness (Falchikov & Goldfinch, 2000; Patri, 2002; Topping, 2009).

Technological advances are further expanding these possibilities. Interactive video platforms and mobile-assisted peer feedback tools support asynchronous, contextualized commenting and allow learners to engage with feedback at their own pace. Studies show that these tools can enhance reflection, autonomy, and affective engagement while improving performance outcomes (Cuocci et al., 2023; Hung, 2019; Zhang & Zou, 2022). Emerging research also suggests that dialogic or reciprocal peer-feedback cycles, where learners respond to each other's comments and clarify intentions, can deepen reflection and foster a sense of shared agency (Gravett & Carless, 2024; Yan & Carless, 2022). Integrating explicit instruction in feedback literacy has been found to not only improve the quality of peer comments but to build learners' confidence in using evaluative information to revise their own speaking in the future (Winstone, Nash, Rowntree, & Parker, 2017). When supported by scaffolds, exemplars, and feedback literacy instruction, peer assessment in speaking becomes not only manageable but a powerful means of promoting self-regulated learning, intercultural awareness, and communicative competence.

In response to these opportunities and challenges, this academic report has two aims: (a) to describe a four-stage framework for embedding peer assessment in tertiary listening–speaking courses in a way that is manageable for both teachers and students, and (b) to share practitioner reflections on how the framework has influenced learner engagement, assessment literacy, and conceptions of effective communication.

2. Practical Framework for Classroom Implementation: Four Stages

Every teaching context is unique, presenting specific challenges as well as opportunities for both the students and teachers to learn. With this in mind, this section presents the pedagogical rationale behind implementing a peer-assessment component and provides a flexible sequence that can be adapted to teachers' own classroom and technological environments. The four stages below describe a practitioner-oriented model derived from the literature and refined through classroom use, rather than a formal research methodology. The activities here are presented as adaptable examples of practice emerging from ongoing teaching experience and have been found by the authors to be both pedagogically grounded and manageable for teachers and students alike. It presents peer assessment in four key stages: (1) collecting benchmark speaking samples, (2) establishing assessment criteria through collaborative ranking, (3) constructing a learner-generated rubric, and (4) implementing peer assessment via technology. Following this sequence can help students develop the assessment literacy needed to evaluate and act on peer feedback.

2.1 Stage 1: Collecting Benchmark Speaking Samples

Stage 1 was used to collect benchmark speaking samples that represented realistic performance by students at approximately CEFR A1–B2 proficiency. In communication-focused English as an International Language (EIL) classrooms, particularly those aimed at developing everyday conversational competence, learners frequently assume that their primary objective is to achieve a form of *nativelike* fluency as modeled by teachers or textbook audio materials. However, research has shown that such expectations can be both unrealistic in the short term and discouraging, as they may imply that non-native varieties of English lack legitimacy (Jenkins, 2007; Seidlhofer, 2011). Failure to recognize learners' own developing English repertoires risks undermining their motivation and communicative confidence. Assumptions about the need to attain nativelike fluency also risk creating a situation whereby students' understandings of assessment and assessment criteria are vague, unrealistic, and overly abstract. Unclear assessment criteria may also result in misalignment with the teacher's expectations for the class.

To counter these traditional second-language-learning-based assumptions, speaking samples can be collected before the course begins from students outside the teacher's immediate class cohort who volunteer to provide examples. Ideally, the speaking samples are videos of the volunteers demonstrating the same tasks that the students will undertake during the upcoming course. Evidence suggests that learners respond positively to listening to speech samples from comparable learners and finding reassurance in their intelligibility and effectiveness (Derwing & Munro, 2015). To maintain comparability, volunteers should share similar language-learning backgrounds with the students, while representing a range of expected proficiency levels. This approach mitigates potential student anxiety about assessment, creates a hierarchy of learning goals, and allays ethical concerns the teacher may have around sharing in-class student work (Borg, 2015).

2.2 Stage 2: Establishing Assessment Criteria through Collaborative Ranking

Stage 2 begins at the outset of the course, when students are informed that they will assess a peer's self-recorded speaking task. Students then participate in three steps: (1) Individually: They rank three to four of the speaking samples collected in Stage 1, relying primarily on intuitive judgments. (2) Group Discussion: They compare individual rankings and are asked to arrive at a shared group ranking. Students are also asked to discuss and note down the criteria behind their rankings. (3) Whole Class Discussion: Groups share their rankings with the whole class as the teacher encourages students to articulate the criteria underlying their decisions. Group rankings can be further aggregated and compared with teacher rankings, opening discussion on what constitutes *effective* communication in EFL contexts.

Eliciting and documenting student language during these discussions allows a set of assessment descriptors to emerge. Common categories may include: Clarity (pronunciation, intonation, rhythm, volume), Fluency (pauses, pace, self-corrections), Language Use (grammatical accuracy, lexical appropriacy), and Content (relevance and supporting details). Depending on cohort dynamics, additional factors such as organization and non-verbal communication may also be discussed and included. The teacher may then collate the criteria generated by the students and the language used to describe them and introduce more formalized labels for the criteria, organizing related

items into coherent categories. This process gradually introduces students to the metalanguage of assessment in a manner that both incorporates and refines their own language use when describing the criteria they used for their rankings, and what they think constitutes effective speaking.

When learners participate in constructing rubrics and assessment criteria in this way, they develop assessment literacy and a deeper understanding of learning goals (Andrade & Du, 2005). This democratization of assessment aligns with Shohamy's (2001) call for shared power and representation in language and has been shown to enhance learner agency and transparency in evaluation (Boud & Falchikov, 2006; Carless, 2007; Topping, 2010).

2.3 Stage 3: Constructing a Learner-Generated Rubric

Following the identification of criteria, the teacher and students work collaboratively to develop a rubric to guide peer assessment. This process begins with a blank rubric template shared with students, which is then completed through whole-class discussion. To ground abstract criteria in observable performance, recorded samples from Step 1 also serve as *living benchmarks*. For instance, as seen in Table 1, the highest-ranked sample informs descriptors of criterion at the highest level, while intermediate samples facilitate differentiation of lower scores. This method aligns with Sadler's (1989) principles of formative assessment, which emphasize concrete examples over vague descriptors. Scaffolding can include banks of feedback phrases collected from students' own language use in earlier group and class discussions, further supporting the development of their assessment metalanguage (Hyland & Hyland, 2006).

Table 1

Excerpt From a Learner-Generated Rubric (Clarity criterion)

Clarity: Pronunciation, intonation, rhythm, volume		
Score	Descriptors	Useful feedback phrases:
3	Easy to understand (Speaking Sample Rank 1)	Positive points: Your words were clear. I could easily understand what you said.
2	Mostly easy to understand (Speaking Sample Rank 2)	
1	Sometimes difficult to understand (Speaking Sample Rank 3)	Points to improve: Sometimes it was hard to catch certain words. Try to say each word a little more clearly.
0	Difficult to understand	

2.4 Stage 4: Facilitating Peer assessment via Technology

Technology-mediated peer assessment has been found to enhance learner engagement and autonomy (Noroozi et al., 2016). In this subsection, the authors describe the minimum procedures required for teachers to replicate a Moodle–H5P workflow. The authors focus on the core steps and decision points without providing guidelines for every possible technical configuration so that teachers can adapt the process to their own institutional settings. To ensure that teachers can feasibly adopt this activity, the authors

include key implementation notes alongside the pedagogical rationale behind each step. H5P Interactive Video (Joubel, 2014), available as an open-source plugin on Moodle (Moodle, 2024) and other learning management systems (LMS), enables students to embed comments directly into their peers' self-recorded speaking performances. This tool allows for structured, dialogic peer assessment that is contextual, repeatable, and closely aligned with assessment criteria.

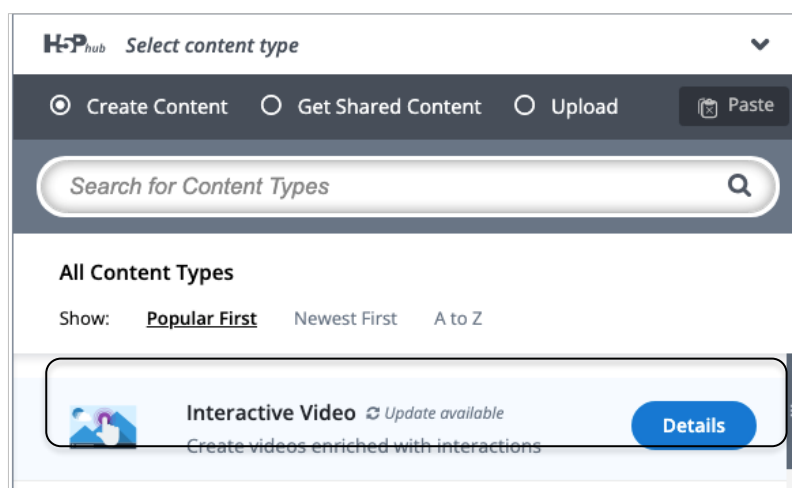
Steps 1 to 5 below illustrate the main steps of this process as supported by the H5P Interactive Video plugin in the Moodle LMS as demonstrated by Student B providing feedback on a video of Student A and vice-versa. The teacher first models the process to the class, demonstrating how to use the software and how to link feedback directly to the rubric criteria and feedback phrases. For example, the teacher can show how to insert a written comment or an audio file at the exact point in a recording where a student mispronounces a word, illustrating how specific feedback aligns with the rubric.

Step 1: Students A and B record a short speaking task for homework that has already been rehearsed in class and mirrors the syllabus aims and objectives and the speaking samples collected in Stage 1 (e.g., presentations, role-plays, or conversations) using smartphones, webcams or tablets. The teacher provides clear guidelines on task length, topic, file format, and basic recording procedures, and students save their recording in a location that can later be uploaded to Moodle.

Step 2: Student A accesses Moodle and adds an H5P Interactive Content activity to the designated course for the peer assessment task, as seen in Figure 1:

Figure 1

Screenshot of an H5P interactive Video Activity Within Moodle Showing the Interactive Video Content Type



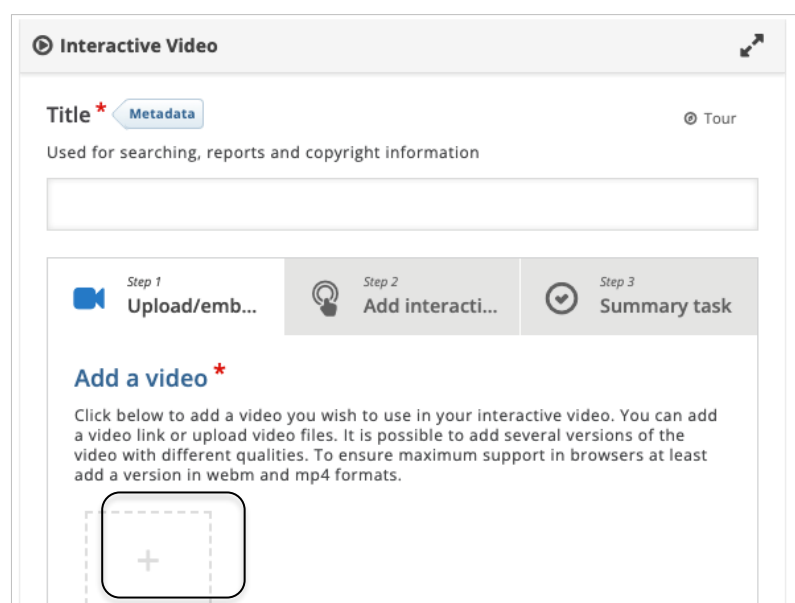
Note. © Moodle Pty Ltd; H5P © Joubel AS, used under open-source licenses. Screenshot by authors.

Step 3: Student A uploads his video to Moodle using the “Add a video” tab inside the H5P Interactive Video activity, as seen in Figure 2, and clicks “Save and return to course.”

Once the video has processed, Student A checks that playback and commenting are functioning correctly and then notifies Student B that her video is ready for peer assessment:

Figure 2

Screenshot of an H5P Interactive Video Activity Within Moodle Showing the Upload Video Interface



Note. © Moodle Pty Ltd; H5P © Joubel AS, used under open-source licenses. Screenshot by authors.

Step 4: Student B opens Student A’s saved activity and selects the “Add interactions” tab (see Figure 3). This function allows Student B to watch Student A’s video and insert interactive feedback such as text pop-ups, audio comments, or reflection prompts using the menu icons displayed at the top of the H5P window (Figure 3). Student B follows the sequence below:

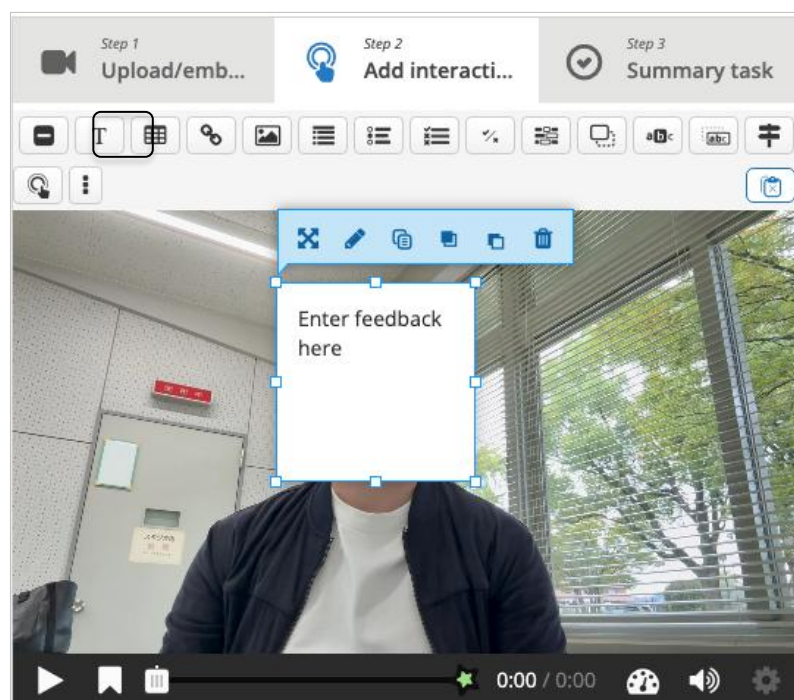
- Before viewing: Skim the rubric criteria and any task instructions so that expectations for content, language, and delivery are clear.
- While viewing: Play the video at normal speed, pausing at key moments to insert time-linked comments that reference specific rubric criteria (for example, clarity, fluency, language use, content).
- After viewing: Re-read the comments, assign scores for each rubric category, and add one or two summary comments that highlight strengths as well as specific, actionable suggestions for improvement.

This structure helps Student B balance holistic impressions of the performance with criterion-referenced judgments. For instance, if Student B has difficulty hearing or understanding what Student A says at a particular moment in the recording, she can click the text icon to open a dialogue box and enter written feedback directly at that point in the video. Drawing on the completed rubric sheet, such as the Clarity criterion illustrated in Table 1, Student B might insert a feedback phrase like: “Sometimes it was hard to catch

certain words. Try to say each word a little more clearly.”

Figure 3

Screenshot of an H5P Interactive Video Activity Within Moodle Showing the “Add Interactions” Tab, Interaction Menu, and a Text Pop-Up



Note. © Moodle Pty Ltd; H5P © Joubel AS, used under open-source licenses. Screenshot by authors.

B adds criteria-based feedback and scores, then clicks “Save.”

Step 5: Student B notifies Student A that feedback has been added to his video and Student A accesses the activity again. As Student A watches, the video will pause at points where Student B’s feedback has been added. Student A is encouraged to record a “targeted redo” of his speaking task based on Student B’s feedback. The cycle may be repeated two or three times. Optionally, the teacher can monitor these iterations and add to Student B’s feedback or provide their own.

2.5 Implementation Notes for Teachers

In the authors’ experience, these steps provide sufficient structure for teachers to implement the activity independently while allowing flexibility for different course designs and institutional Moodle settings. To support teachers wishing to adopt this activity, this subsection outlines the minimum practical requirements for implementing the Moodle–H5P workflow. First, teachers must ensure that students have permission to create or edit H5P activities within the course; this setting is typically controlled by a course administrator and may require a one-time adjustment. Second, the teacher creates a designated section in the Moodle course (e.g., “Peer Assessment Tasks”) and models the full workflow once, including how to add an H5P activity, upload a video file, and

insert time-linked comments. During modeling, teachers should also demonstrate how to align comments with rubric criteria, how to save interactions, and how to notify a peer when the task is complete. Third, students require access to a device capable of recording short videos (smartphones, tablets, or webcams) and a stable means of uploading those files. Finally, the teacher should schedule the activity so that students have clear deadlines for (a) uploading their recording, (b) adding peer feedback, and (c) submitting their “targeted redo.” These procedural elements enable teachers to implement the activity smoothly, even if they are new to H5P or technology-mediated peer assessment.

Because Steps 1 to 5 are asynchronous in nature, they allow both peer assessors and peer-assessed students to give and receive feedback in a contextualized, reflective, and self-paced manner. This flexibility enables students to engage more deeply with the assessment task, review their peers’ performances without the pressure of real-time interaction, and consider their comments more thoughtfully. Conditions like these are conducive to enhanced feedback quality and uptake, as students have time to rewatch, pause, and reflect before providing their feedback or acting upon it (Carless & Boud, 2018; Nicol, 2020).

This asynchronous design also provides an opportunity for a degree of anonymization, which can improve reliability and fairness by reducing the social biases often associated with face-to-face peer assessment (Falchikov & Goldfinch, 2000; Patri, 2002; Topping, 2009). Although peer assessment of speaking, particularly with the use of video, does not allow the speaker complete anonymity, an anonymous peer-assessor might help mitigate issues such as friendship bias, reciprocity pressures, and reluctance to offer critical feedback, leading to more objective and honest evaluations. Matching each student with a different peer reduces the likelihood of reciprocal leniency and reinforces inter-rater consistency across the cohort.

It must be noted that introducing anonymized feedback entails practical and administrative demands, as teachers must design, monitor, and maintain systems that preserve anonymity while ensuring accountability and pedagogical coherence. Furthermore, without interaction between peer-assessors themselves, opportunities for clarification and dialogue are limited, which may result in misunderstandings or surface-level engagement with feedback. Moreover, anonymity can reduce accountability if students perceive feedback they give as detached from personal responsibility.

These risks can be mitigated through structured reflection tasks and feedback literacy training that help learners interpret and act on feedback constructively while maintaining fairness and trust (Carless, 2020; Winstone et al., 2017). One practical example of mitigating these risks is with structured reflection tasks. After receiving peer assessment, students can complete a reflection sheet designed to guide them through interpreting and responding to the feedback constructively. They begin by identifying one peer comment they agree with and another they find unclear or potentially unfounded. Next, they paraphrase each comment in their own words to ensure comprehension and then set one specific action goal for improvement, such as slowing their speech rate or emphasizing key words in future recordings. If clarity is requested a teacher can act as intermediary between the peer-assessor and the peer-assessed. This reflection task can be utilized even when anonymization of peer assessment is not undertaken. Overall, anonymization requires careful evaluation of its potential to enhance fairness and objectivity against the practical and pedagogical challenges it may introduce in a specific teacher’s classroom context.

Although the framework described in Section 2.4 is illustrated using Moodle–H5P Interactive Video, the pedagogical logic of the activity is not platform-dependent. At a minimum, the technology must allow students to (a) record short speaking performances, (b) access peer recordings, and (c) provide time-referenced or performance-specific feedback. In practice, this can be achieved with relatively modest technical resources. For the Moodle–H5P workflow, minimum requirements include student access to a personal recording device (smartphone, tablet, or webcam), an LMS that permits video upload and playback, and permission for students to create or edit H5P Interactive Video activities. Apart from Moodle, H5P interactive software is available as a free plugin on platforms such as Canvas, Brightspace, Blackboard, WordPress, and Drupal. No specialist software beyond a web browser is required, and videos can be recorded using default camera applications on student devices.

Where H5P is unavailable, teachers can adopt lower-tech or hybrid alternatives while preserving the core feedback cycle. One option is a paper-based approach in which students watch peer videos together in class or independently, complete a printed rubric, and deliver feedback face-to-face or via short response videos of their own. A semi-digital alternative involves adding subtitles or captions to videos as a form of feedback. For example, students can insert comments directly into subtitle files by typing feedback into a Word document with pre-prepared timestamps and then saving it as a plain-text SubRip (.srt) file. If the filename of the .srt file is the same as the video file, feedback (in the form of subtitles) will play automatically. This method allows feedback to appear synchronized with specific moments in the performance without requiring interactive video software. Similarly, widely available platforms such as YouTube provide built-in captioning and subtitle tools that can be repurposed for peer feedback, enabling students to add time-aligned written comments privately or in unlisted videos. Comparable captioning or annotation functions are also available in other common video-editing or presentation tools that might be more familiar to teachers. Across all variants, the essential pedagogical requirement remains the same: feedback must be specific, criterion-referenced, and linked to observable moments in the speaking performance, followed by an opportunity for revision or targeted re-recording, as outlined by Black and Wiliam (2009).

3. Pedagogical Implications and Practitioner Reflections

The observations in this section illustrate how the framework has functioned in the authors' classrooms and how these experiences align with themes in the wider literature. Definitive claims would require systematic data collection and rigorous analysis, so the reflections offered here should be regarded as tentative and situated within a particular classroom context. What follows draws on repeated implementation of the framework across multiple cohorts in required first-year English communication courses and on teachers' classroom observations and informal student feedback.

3.1 Illustrative Outcomes

Across several semesters, the authors have observed that when students take on the role of assessors, the authors notice clear gains in autonomy and engagement. This shift encourages students to participate more actively and reflect more deeply, an observation supported by research into formative peer assessment and learner agency (Carless, 2007; Nicol & Macfarlane-Dick, 2006). Their students also show a greater ability to self-assess and revise their speaking once they are asked to act on peer comments. This reflects

Sadler's (1989) principle that feedback is most powerful when it enables learners to close the gap between current and desired performance. In practice, the authors have found that their students are more motivated to re-record their speaking tasks when feedback is concrete, contextualized, and tied to observable performance.

The authors also observe that their students begin to redirect their goals away from unattainable native-like ideals and toward intelligibility and mutual understanding. This mirrors similar findings in EIL research about a shift to mutual intelligibility, rather than a native-speaker model (Derwing & Munro, 2015; Jenkins, 2007; Seidlhofer, 2011), and is evident when students emphasize intelligibility in post-task reflections.

Another outcome is the development of a shared classroom discourse around performance. By reusing the language that their students generate during assessment discussions together, the authors establish a collective metalanguage that makes feedback more precise and accessible, consistent with Hyland and Hyland's (2006) observations. The authors have noticed gradual improvements in the specificity of student feedback. Early in the course, peer comments are often global and impressionistic (for example, "good pronunciation" or "speak more clearly"). After working with learner-generated rubrics and seeing multiple annotated exemplars, students are more likely to anchor comments in observable behavior (for example, "Your eye contact dropped when you looked at the slides" or "The pause before your key point made it easier to follow"). Over time, some learners begin to anticipate these criteria while planning and rehearsing their own speaking tasks, suggesting a shift toward more self-regulated performance.

These impressions are based on informal classroom observations and end-of-course reflections rather than systematic data collection, and should therefore be read as local, exploratory insights rather than generalizable findings.

3.2 Challenges and Design Considerations

At the same time, the peer-assessment framework presented in this academic report introduces important affective and relational challenges. There are certainly students who struggle to articulate assessment criteria, who feel the emotional weight of critiquing classmates, and/or who hesitate to adapt to new technologies. Some students worry about "not being good enough" at English to evaluate peers and may initially default to overly generous scores in order to avoid potential conflict.

There are also practical considerations related to time and cognitive load. Training students to use H5P Interactive Video, to apply the rubric consistently, and to compose helpful comments requires multiple cycles of modeling and guided practice. Without sufficient support, some learners focus on the mechanics of the software rather than the content of their feedback, which can dilute the formative value of the activity.

Finally, decisions about anonymity and accountability need careful thought. Anonymizing feedback can reduce anxiety for both assessors and recipients, but it may also weaken students' sense of responsibility for the comments they provide. In their context, the authors found it more productive to keep identities visible while explicitly discussing expectations for respectful, constructive language and modeling how to phrase critical feedback in supportive ways.

Based on these experiences, the authors offer the following design recommendations for teachers considering adoption of this framework:

- Allocate at least one full class to modeling the workflow and feedback language, including sample comments that balance praise and critique.

- Use short, low-stakes speaking tasks for the first peer-assessment cycle so that students can focus on learning the process rather than worrying about high-stakes grades.
- Provide sentence stems and discourse moves (for example, “One thing that helped me understand was...,” “A specific way to improve might be...”) to support both positive and critical feedback.
- Decide early whether anonymization is feasible and pedagogically appropriate and make the rationale explicit to students.

Overall, these challenges highlight the importance of designing peer-assessment processes that are transparent, supportive, and developmentally appropriate for learners with diverse linguistic and technological backgrounds. What emerges is a shift from a teacher-centered instruction toward a more collaborative and reflective classroom community supported by authentic samples, co-constructed rubrics, and iterative cycles of peer feedback.

4. Conclusion

This academic report has offered a situated account of one implementation and the practitioner insights that emerged from it. While the peer-assessment framework outlined here draws on established research in formative assessment and learner autonomy, the emphasis here is on practical classroom procedures. The framework lends itself to a range of required and elective classes emphasizing listening and speaking skills; it can be adapted to a range of EFL settings to align with specific institutional goals, technological resources, and learner profiles. Future research might build on this framework by systematically examining how different configurations of benchmarking, rubric co-construction, and technology-mediated commenting influence learners’ feedback literacy, speaking performance, and beliefs about assessment. It is hoped that this kind of research would confirm the authors’ experience that feedback is a transferable skill: even after the course has ended, when no teacher or textbook is available, learners can continue to improve their speaking through the habits of self and peer assessment they have developed in the course.

References

- Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research, and Evaluation*, 10(3).
- Bennett, R. E. (2024). Formative assessment and the future of educational measurement. *Educational Measurement: Issues & Practice*, 43(1), 8–23.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.
- Borg, S. (2015). *Teacher cognition and language education: Research and practice*. Bloomsbury.
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399–413.
- Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400–413.
- Carless, D. (2007). Learning-oriented assessment: conceptual bases and practical

- implications. *Innovations in Education and Teaching International*, 44(1), 57-66.
- Carless, D. (2015). *Excellence in university assessment: Learning from award-winning practice*. Routledge.
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325.
- Carless, D., & Winstone, N. (2020). Teacher feedback literacy and its role in student learning. *Review of Education*, 8(2), 521–541.
- Cuocci, S., Marnani, P. F., Khan, I., & Roberts, S. (2023). A meta-synthesis of technology-supported peer feedback in ESL/EFL writing classes. *Languages*, 8(2), 114.
- Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. John Benjamins.
- Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, 32(2), 481–509.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287–322.
- Frey, N., Fisher, D., & Hattie, J. (2018). Developing “assessment capable” learners. *Educational Leadership*, 75(5), 46–51.
- Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education*, 36(2), 137–155.
- Gravett, K., & Carless, D. (2024). Feedback literacy-as-event: relationality, space and temporality in feedback encounters. *Assessment & Evaluation in Higher Education*, 49(2), 142-153.
- Hyland, F., & Hyland, K. (2006). Feedback on second language students’ writing. *Language Teaching*, 39(2), 83–101.
- Hung, Y. (2019). Enhancing peer feedback through digital platforms in EFL speaking classes. *Computer Assisted Language Learning*, 32(7), 680–704.
- Jenkins, J. (2007). *English as a Lingua Franca: Attitude and Identity*. Oxford University Press.
- Joubel AS. (2014). *H5P (Version 1.12) [Computer software]*. <https://h5p.org>
- Liu, N.-F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education*, 11(3), 279–290.
- Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.
- McNamara, T. (1996). *Measuring second language performance*. Longman.
- Moodle Pty Ltd. (2024). *Moodle (Version 5.1.1) [Learning management system software]*. <https://moodle.org>
- Nicol, D. (2020). Re-imagining feedback in higher education: Developing students’ evaluative judgement. *Higher Education Research & Development*, 39(1), 4–17.
- Nicol, D. J., & 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.
- Noroozi, O., Biemans, H. J., & Mulder, M. (2016). Relations between scripted online peer feedback processes and quality of written argumentative essay. *The Internet and Higher Education*, 31, 20–31.
- Patri, M. (2002). The influence of peer feedback on self- and peer-assessment of oral

- presentation tasks. *Assessment & Evaluation in Higher Education*, 27(2), 135–144.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15(1), 1–20.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144.
- Saito, H. (2008). EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing*, 25(4), 553–581.
- Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford University Press.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Pearson Education.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312.
- Topping, K. J. (2009). Peer assessment. *Theory Into Practice*, 48(1), 20–27.
- Topping, K. J. (2010). Peers as a source of formative assessment. In J. H. McMillan (Ed.), *Formative Classroom Assessment: Theory into Practice* (pp. 61–74). Teachers College Press.
- Topping, K. J. (2018). Peer assessment: Learning by judging and discussing the work of other learners. *Interdisciplinary Education and Psychology*, 2(1), 1–17.
- Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, 52(1), 17–37.
- Yan, Z., & Carless, D. (2022). Self-assessment is about more than self: the enabling role of feedback literacy. *Assessment & Evaluation in Higher Education*, 47(7), 1116–1128.
- Zhang, R., & Zou, D. (2022). A review of research on technology-enhanced peer feedback for second language writing. *Education and Information Technologies*, 28, 6727–6753.