

Interactive Podcast-Based Learning Model to Enhance Speaking Skills of University Students

Rio Kurniawan^{1*}, Azhar Aziz Lubis², Panut Setiono³, Didi Yulistio⁴

¹⁻⁴ Universitas Bengkulu, Indonesia

Abstract A design by which every learner is given repeated opportunities to speak, personal performance is reviewed, actionable feedback is obtained, and a subsequent attempt is improved is needed for university-level speaking instruction. In this study, a platform-neutral interactive podcast learning sequence for third-semester students in an Indonesian Language Education programme at Universitas Bengkulu was developed and preliminarily evaluated. A development-focused mixed-methods approach, which is expressed as QUAL → quan, was integrated into the Analysis, Design, Development, Implementation, and Evaluation stages of ADDIE. The formal needs analysis was drawn on three classroom observations, interviews with 15 students, two lecturers, and one educational-technology specialist, as well as a course-document review. Five podcast cycles over ten meetings were completed by thirty students from one intact class. Following revision, the expert-appraisal score was raised from 85.0% to 88.5% of the maximum score. Strong practicality was reported by the students ($M = 4.18$, $SD = 0.46$), while a combined descriptive mean of 4.38 was produced by the two lecturers. The composite speaking-performance score, which is based on fluency, pronunciation, and intonation, was increased from 63.33 ($SD = 8.10$) to 81.00 ($SD = 6.70$), $t(29) = 12.12$, $p < .001$, $d_z = 2.21$. Observable speaking assurance was also improved, although it was analysed separately from linguistic performance. Because a single intact class was used, no comparison group was included, and detailed process measures for feedback quality and individual learning trajectories were lacked by the evaluation, feasibility and preliminary promise, rather than causal effectiveness, are supported by the findings. The principal contribution is a transferable practice–replay–feedback–revision sequence, which is supported by culturally responsive task criteria, restricted sharing, and accessible technical alternatives.

ARTICLE HISTORY

Received : 30 March 2026

Revised : 4 July 2026

Accepted : 12 July 2026

KEYWORDS

Feedback literacy; Instructional design; Interactive podcast; Learner-generated audio; Speaking skills

PUBLISHER'S NOTE

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) license.



CORRESPONDING AUTHOR

*Rio Kurniawan, Universitas Bengkulu, Indonesia. Email: kurniawanrio@unib.ac.id

Introduction

Speaking competence is central to the professional preparation of future language teachers. They are required by classroom responsibilities to explain concepts, ask purposeful questions, tell stories, provide feedback, and adjust communication to different audiences. It is indicated by research on technology-enhanced language learning that oral development is more likely when language is repeatedly produced by learners, their own performances are revisited, and explicit criteria are responded to, rather than technology being used only as a channel for the consumption of content (Huang & Li, 2024; Perry, 2024). In this context, student-generated podcasts may be pedagogically useful, but their value is depended on how recording, reflection, feedback, and revision are organised (Wakefield et al., 2023).

The project was begun with informal scoping in May 2025, followed by a formal needs assessment in July 2025. The early visits were served only so that a possible instructional difficulty was located, and they were not included as research evidence. Formal evidence was obtained from three structured observations in the selected 30-student class, interviews with 15 students,

interviews with two lecturers and one educational-technology specialist, and an analysis of the course syllabus and the Semester Learning Plan. The kinds of speaking opportunities and feedback that were available were described by the observation records, but the speaking-turn duration and the feedback frequency were not measured by them. Consequently, the baseline is interpreted as a set of convergent design requirements rather than as a quantified prevalence.

The core difficulty was not the complete absence of oral work. Classroom presentations were already made by the students; yet, the opportunities for individual speaking were uneven, class time was dominated by a limited number of performances, and feedback was often broad rather than tied to specific criteria. Few opportunities to listen to their own voices and to revise their work before the final assessment were also had by the students. Inconsistent device quality and concern about public distribution were further identified by the interview and observation evidence. The instructional problem was therefore defined as the need for an accessible and privacy-conscious sequence, by which speaking practice is distributed across the class and self-review, peer interaction, lecturer response, and revision are connected.

The design was informed by four principles. First, it is ensured by repeated individual recording that oral production is not concentrated among only the most confident students. Second, pauses, articulation, pronunciation, and intonation are made available for deliberate inspection by replay. Third, students are treated by feedback literacy as active users of criteria, by whom evaluative information is learned to be produced, interpreted, and applied. Peer feedback is more useful when guidance is received by learners and when revision is built into the task rather than being left optional (Carless & Boud, 2018; Little et al., 2024; Man et al., 2024). It is similarly shown by evidence from mobile-assisted speaking activities that feedback quality is depended on motivation, scaffolding, and the interaction design (Xu & Peng, 2022). Fourth, learners' cultural knowledge is valued by culturally responsive digital pedagogy, while choice, accuracy, respectful representation, and alignment with disciplinary outcomes are required (Gay, 2018; McDaniel, 2024; Wesley-Nero & Donley, 2024).

Several relevant benefits are reported by existing studies, but they are not always integrated into one instructional cycle. Learner-generated podcasts may be served as authentic assessment, and the pressure that is associated with live speaking may be reduced (Perry, 2024; Wakefield et al., 2023). Oral proficiency and the willingness to communicate can be supported by digital storytelling when meaningful narratives are constructed by students (Huang, 2023), and communication barriers may be reduced by technology-enhanced environments when pedagogy, privacy, and learner diversity are considered (Huang & Li, 2024). Nevertheless, the final product is often emphasised by podcast assignments, mandatory uptake may be omitted by peer-feedback studies, and culturally situated tasks may not be connected to criterion-referenced oral revision. These elements are integrated into a testable instructional proposition by the present study.

Table 1. Comparative synthesis of related instructional approaches

Approach	Typical emphasis	Limitation for the present problem	Distinctive feature in the proposed sequence
Teacher-produced podcast	Listening input and flexible access	Students may remain consumers and complete no speaking-revision cycle	Each learner records and revises individual audio

Learner-generated podcast	Planning, recording, and sharing a product	Feedback uptake may be optional or focused on product quality	Peer and lecturer feedback is criterion-referenced and followed by revision
Mobile-assisted peer feedback	Digital exchange of comments on oral work	Quality depends on training, motivation, and task structure	Three-prompt scaffold, moderation, and a lecturer-response rule
Digital storytelling	Meaningful narrative production and identity work	May not target specified oral-performance criteria	Local narratives are linked to speaking objectives and ethical representation
Proposed sequence	Practice → replay → feedback → revision	Requires moderation and technical support	Platform-neutral workflow integrating self-monitoring, feedback literacy, cultural responsiveness, and restricted sharing

Figure 1. Conceptual architecture of the platform-neutral podcast sequence

Design foundation	Operational sequence	Implementation supports	Evidence examination
Repeated individual practice; replay and self-monitoring; feedback literacy; culturally responsive design	Brief and topic choice; planning → recording/replay/self-assessment → feedback → lecturer feedback → revision/reflection	Restricted sharing; respectful moderation; possible recording; lecturer response rule	Expert feasibility; implementation; speaking performance; observation

The product was defined as an instructional sequence rather than as a particular podcast application. Task syntax, student and lecturer guidance, speaking criteria, feedback prompts, moderation procedures, privacy rules, and revision requirements were included within its boundaries. The applications were replaceable. During the trial, available audio recorders, restricted cloud links, and a moderated messaging group were used by the students. Public podcast release and Spotify distribution were not made part of the tested intervention.

Cultural responsiveness was translated into five operational criteria: alignment with the Speaking Skills objective; the use of local narratives as learning resources rather than stereotypes; student choice in the selection or proposal of a suitable topic; lecturer review for accuracy and respectful representation; and reflection on audience, language choice, and ethical retelling. Lecturer consultation and course alignment were confirmed by the available records, but a systematic audit of topic changes or cultural-accuracy decisions was not contained by them. These elements were therefore retained as requirements for future implementation rather than being reported as fully verified outcomes.

Five questions were addressed by the study: (1) Which needs and design requirements were emerged from the formal assessment? (2) How was the sequence developed and revised through ADDIE? (3) What was indicated about feasibility by the expert appraisal, the practicality responses, and the implementation records? (4) What preliminary change occurred in the three-component speaking score? and (5) What preliminary change occurred in observable speaking assurance when it was treated separately from linguistic performance?

Method

Design and Mixed-Methods Integration

In this research-and-development study, ADDIE (Analysis, Design, Development, Implementation, and Evaluation) was used as the product-development framework (Branch, 2009). The methodological sequence is represented as QUAL → quan: priority was given to the qualitative needs evidence in Analysis and Design; descriptive quantitative evidence was embedded in Development and Implementation; and pretest–posttest statistics were used in Evaluation. Integration was occurred through connecting, building, and joint interpretation. Qualitative themes were connected to the design requirements, revisions were built by the appraisal and pilot evidence, and the interpretation of quantitative change was delimited by the implementation evidence (Creswell & Plano Clark, 2018). The inference that was generated by the integration was not that improvement was caused by the sequence, but that the design was feasible in one cohort and sufficiently promising for controlled evaluation.

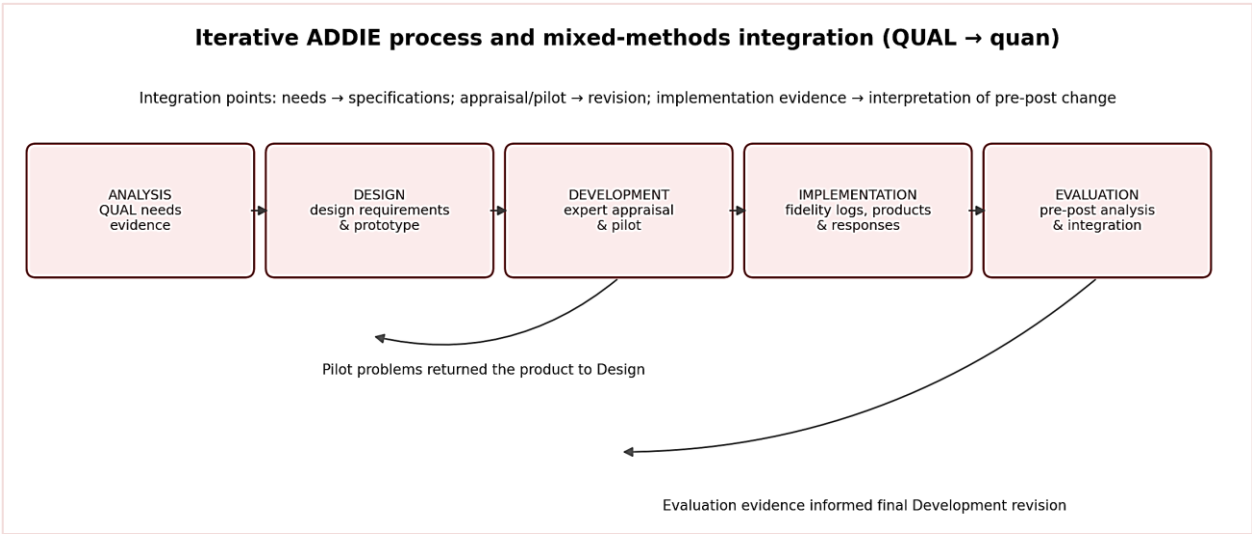


Figure 2. Iterative ADDIE process, decision points, and mixed-methods integration

Figure 2 illustrates the iterative application of the ADDIE framework within a qualitatively driven mixed-methods design. The Analysis, Design, Development, Implementation, and Evaluation stages are systematically connected through multiple integration and feedback points. Revisions to the initial design are informed by the evidence that is obtained from expert appraisal and pilot testing, while interpretation and final product refinement are guided by the implementation and pre–post evaluation data. It is ensured by this cyclical process that development decisions remain evidence-based, responsive, and methodologically integrated.

Table 2. Alignment of research questions, evidence, analysis, and results

Research question	Primary evidence	Analysis	Result location
RQ1: Needs and requirements	Three observations; 15 student interviews; 2 lecturer interviews; 1 technology-specialist interview; syllabus/RPS	Cross-source coding and design-requirement matrix	Needs assessment and design requirements
RQ2: Development and revision	Design records, expert comments, pilot logs	Audit trail with decision triggers and iteration records	Development iterations and expert appraisal
RQ3: Feasibility	Expert appraisal, practicality questionnaire, implementation logs	Percentage of maximum descriptions; means/SD for students; descriptive lecturer summary; fidelity evidence table	Feasibility, practicality, and fidelity
RQ4: Speaking-performance change	Coded pretest/posttest audio ratings	Paired t test, 95% CI, dz; exploratory dimension tests with Holm threshold	Preliminary learning change
RQ5: Observable assurance change	Separate rubric indicator	Paired t test and cautious descriptive interpretation	Preliminary learning change

Development Procedure and Decision Rules

The procedure was iterative rather than strictly linear. The first specification was generated by the needs evidence. During Development, the product was returned to Design by the validator and pilot evidence: restricted sharing was led to by privacy concerns, ambiguous rubric descriptors were rewritten, and the peer-feedback template was reduced to three prompts. During Evaluation, the product was returned to Development for final clarification by inconsistencies in the score conversion and interpretation. A revision was triggered when at least one of the following was occurred: the same conceptual problem was identified by two or more validators; a required step could not be completed by a pilot participant without assistance; a privacy or respectful-interaction risk was identified; or it was shown by the implementation logs that a procedure was not feasible with the available devices or connectivity.

Table 3. Development audit trail and documented iteration

Stage and evidence	Decision trigger	Revision/iteration	Output
Analysis (July 2025): observations, interviews, documents	Repeated evidence of limited individual practice, general feedback, device inequality, and privacy concern	Specified individual recording, replay, criterion-based feedback, restricted sharing, and technical alternatives	Needs-to-design matrix

Stage and evidence	Decision trigger	Revision/iteration	Output
Design (August 2025): initial syntax, guides, tasks, rubric	Mismatch between public podcast conventions and participant privacy concerns	Returned to Design; public publication removed and platform-neutral functions specified	Prototype sequence and guides
Development (September 2025): three-expert appraisal and 10-student pilot	Repeated comments on instruction clarity, privacy, rubric ambiguity, file naming, and low bandwidth	Returned to Design; instructions shortened, privacy rules added, descriptors sharpened, feedback prompts reduced	Revised implementation package
Implementation (October–November 2025): ten meetings	Internet disruption and inadequate microphones for some students	Operational adaptation without changing core pedagogy: compressed audio, shared microphones, campus facilities	Implementation logs and completed recordings
Evaluation (December 2025): appraisal, practicality, pretest/posttest evidence	Need to separate linguistic performance from observable assurance and avoid causal claims	Returned to Development; outcome reporting, limitations, and final specification revised	Final instructional sequence and manuscript

Participants, Sampling, and Research Ethics

The population and the accessible population were each comprised of 60 third-semester students who were registered in two Speaking Skills classes in the Indonesian Language Education Study Program at FKIP Universitas Bengkulu during 2025/2026. One intact class of 30 students was became the intervention cohort, because its schedule was matched with the development period and the integration of the sequence into regular instruction was agreed to by its lecturer. All eligible students in that class participated. Selection bias is created and transferability is limited by this convenience-based intact-class selection. A quasi-experimental comparison could theoretically have been supported by the second class, but synchronised baseline and intervention procedures were prevented by the different scheduling and instructional timing; consequently, the project was treated as a developmental feasibility trial rather than as a comparative evaluation.

The 15 interviewees were purposively selected from the intervention class, so that lower, middle, and higher preliminary speaking levels were represented. Because both the needs assessment and the implementation were contributed to by the same students, expectation and social-desirability effects are possible. The interviews were continued to 15 participants, but a formal saturation grid and a negative-case log were not retained; therefore, no strong saturation claim is made by the revision. Dissenting or technically difficult experiences are reported where available, including privacy concerns, unstable internet, and inadequate microphone quality. An explanation of the activities was received by the students, and consent to voice recording and restricted class use was given. The recordings were shared only through a closed class space, and the students were instructed that peers' files were not to be redistributed.

Instructional Sequence, Dosage, and Platform-Neutral Specification

The sequence was consisted of five cycles, which were implemented across ten regular class meetings. Briefing, planning, recording, replay/self-review, peer feedback, lecturer feedback, and

revision/reflection were included in each cycle. The audio products lasted 3–4 minutes, except the 1–2-minute orientation recording. The duration was selected, so that repeated individual production was made feasible within one course while the files were kept small enough for low-bandwidth submission. The exact out-of-class preparation time and the lecturer workload were not logged, so workload feasibility is interpreted from the completion records and the practicality responses rather than from a time-on-task measurement.

Table 4. Core functions, implementation example, and cycle objectives

Core function / cycle	Platform-neutral requirement	Implementation example in the trial	Learning objective / evidence
Restricted capture and sharing	Any audio recorder and access-controlled sharing space	Student devices/Audacity; restricted cloud links	Protect privacy while enabling replay
Cycle 1: orientation and voice awareness	1–2-minute diagnostic recording and rubric orientation	Trial recording during Meetings 1–2	Recognize voice quality and recording procedure
Cycle 2: pronunciation-focused monologue	3–4-minute individual monologue, replay, criteria-based feedback	Meetings 3–4	Increase attention to articulation and fluency
Cycle 3: culturally grounded narrative retelling	Narrative selected under relevance, choice, accuracy, and respect criteria	Meetings 5–6	Develop expressive retelling and audience awareness
Cycle 4: teaching simulation	Explain a narrative text to an imagined secondary audience	Meetings 7–8	Practice teacher-oriented oral explanation
Cycle 5: reflective final performance	Final recording, post-task reflection, and posttest	Meetings 9–10	Integrate performance criteria and reflection
Feedback and revision	Structured peer prompt, lecturer rubric feedback, opportunity to revise	Moderated class messaging; lecturer response target within 48 hours	Develop feedback literacy and performance uptake

Automated speech scoring, AI-generated feedback, and generative-AI tools were not provided by the study. Scripts could be planned and basic recording functions could be used by the students, but the possible use of external editing, pronunciation, or AI assistance outside the formal workflow was not systematically audited. This boundary is acknowledged, because performance could be influenced by uncontrolled external assistance.

Instruments and Measurement

A 12-question semi-structured interview guide, a 16-indicator observation sheet, and a document-analysis guide were included among the needs-assessment instruments. Speaking opportunity, feedback experience, hesitation, technology access, attitudes toward recording, and cultural-task relevance were covered by the domains. The original instrument package was not included in the manuscript files that were supplied for revision; accordingly, the instrument domains and the analytic use are reported by the article rather than the unverified item wording being reproduced.

A five-point response scale was used by the 18-item practicality questionnaire. Four item clusters—guide clarity, technological ease, feedback usefulness, and learning/cultural relevance—were used, so that descriptive reporting was organised only; it is not claimed by the study that validated latent subscales are constituted by these clusters. A Cronbach's alpha of .89 was had by the total student scale. Item-level distributions and cluster-specific reliability were not available, so the result is interpreted as a descriptive usability indicator that is subject to acquiescence and course-context social-desirability bias. The lecturer responses are presented descriptively, because only two lecturers participated.

Speaking performance was assessed with parallel 3–4-minute teaching-simulation tasks, by which the same genre, audience, duration, and rating criteria but different narrative topics were used. The exact prompt wording and a formal linguistic-difficulty equivalence analysis were not available; only a limited comparability check was provided by the expert review. The primary measured construct was a three-component speaking-performance score, which is comprised of fluency, pronunciation, and intonation. Content organisation and audience awareness were instructional aims, but they were not separately scored, by which the construct is narrowed and which is acknowledged as a measurement limitation. The former “self-confidence” indicator was renamed observable speaking assurance, because it was inferred from performance behaviour rather than from a validated affective scale.

Each dimension was scored from 1 to 5 and was converted to a 0–100 scale. Coded files without pretest/posttest labels were independently rated by two trained speaking-course lecturers. The raters were not fully independent of the instructional context, by which possible expectancy bias is created. The reported $ICC(2,2) = .87$, 95% CI [.78, .93], is interpreted as a two-way random-effects, absolute-agreement, average-measures coefficient (Koo & Li, 2016). Dimension-level ICC values were not retained, so reliability is established only for the reported composite and should be strengthened through external blinded raters in future work.

Expert Appraisal, Qualitative Analysis, and Statistics

The prototype was appraised by three experts: two language-education lecturers with speaking-assessment expertise and one educational-technology lecturer with digital-learning expertise. They were selected for domain relevance; their institutional independence from the research team is not documented by the available records. 26 items on a 1–5 scale were used by the appraisal, and it is reported as the percentage of the maximum obtainable score. No external empirical cut-off was applied, and the percentages are not described as proof of content validity. Item-level agreement, validator-specific totals, and uncertainty estimates were not available; therefore, the results are framed as a limited expert-feasibility appraisal. A second appraisal after revision is indicated by the available record, but blinding to the earlier scores is not documented, so confirmation effects cannot be excluded (Polit et al., 2007; Yusoff, 2019).

Coding, categorisation, and cross-source comparison were followed by the qualitative analysis (Miles et al., 2014). Meaningful interview statements, observed events, and relevant document segments were coded, grouped into need categories, and translated into design requirements. Observation and document evidence were used, so that interview claims were corroborated or qualified, rather than numerical weights being assigned. The preliminary categories were reviewed by a peer debriefer. Because verbatim transcripts, a formal codebook, a saturation table, and a negative-case register were not part of the supplied archive, evidence-source matrices

are reported by the revision, and the claims are limited to development-oriented requirements rather than rich interpretive themes.

In the primary quantitative analysis, the three-component speaking-performance score was compared using a paired-samples t test, a 95% confidence interval, and dz. Fluency, pronunciation, intonation, and observable assurance were treated as exploratory analyses; a Holm-adjusted familywise threshold was applied, and all reported p values below .001 remained below the adjusted criteria. A Shapiro–Wilk $W = .96$, $p = .31$ for the change score was reported by the original analysis. Given $n = 30$ and the very large standardised changes, the effect sizes are treated as provisional. Individual trajectory plots, robust sensitivity analysis, regression-to-the-mean checks, and a definitive ceiling-effect assessment were not permitted by the aggregate outputs (Lakens, 2013).

Results and Discussion

Needs Assessment and Design Requirements

Five cross-source requirements were generated by the formal needs assessment. The strongest convergence was concerned with the limited opportunities for every student to speak individually and with the need for specific, usable feedback. Private rehearsal and replay before public performance were also valued by the students. Technology access was available, but the device quality was uneven. Local narratives were considered useful when they were connected to the course objectives and were represented respectfully. Because verified verbatim quotations or coded prevalence counts were not included in the retained records, Table 5 reports source convergence and observable examples rather than invented participant quotations.

Table 5. Cross-source evidence and resulting design requirements

Need / barrier	Evidence convergence	Observable or documentary example	Design requirement
Limited individual practice	Observations + student/lecturer interviews	Conventional presentations concentrated oral production in relatively few classroom turns	Five cycles requiring an individual recording from every student
General or delayed feedback	Student and lecturer interviews + observation sheet	Comments were often broad rather than tied to pronunciation, fluency, or intonation criteria	Three-prompt peer feedback plus lecturer rubric response
Fear of direct judgment	Student interviews + hesitant turns in observations	Students requested an opportunity to rehearse and listen before presenting	Restricted sharing, replay, and re-recording
Unequal device quality	Technology-access interview questions + implementation logs	Some devices had inadequate microphone quality; internet disruption occurred	Shared microphones, campus facilities, compressed/low-bandwidth submission
Need for meaningful content	RPS/syllabus + lecturer interviews	Narrative and teaching-simulation tasks aligned with Indonesian-language teaching objectives	Culturally responsive topic criteria and student topic agency

Development Iterations and Expert Appraisal

85.0% of the maximum possible score was yielded by the first appraisal, and four recurring revision areas were identified: alignment between tasks and outcomes, usability and privacy of the workflow, clarity of instructions, and precision of assessment descriptors. After revision, the overall value was 88.5%. The improvement is descriptive and should not be interpreted as robust content validity, because only three experts participated and validator-specific variation was unavailable.

Table 6. Expert appraisal by aspect and documented revision

Aspect (items)	Initial % of maximum	Revision	Post-revision % of maximum
Course-content alignment (8)	86.7	Clarified links between each podcast task and course outcomes	90.0
Media/workflow usability (7)	83.8	Simplified procedures; added restricted sharing and privacy safeguards	88.6
Language clarity (6)	86.1	Removed ambiguous wording and added procedural examples	91.7
Assessment alignment (5)	83.5	Separated pronunciation and intonation descriptors; clarified score conversion	87.5
Overall (26)	85.0	Minor but substantive workflow and assessment revision	88.5

The 10-student pilot was conducted with students from the same study program, by whom membership of the 30-student intervention class was not held. Recording, upload, restricted sharing, and comment procedures, rather than learning outcomes, were tested by the pilot. Standardised file naming, a three-prompt feedback template, and a low-bandwidth submission alternative were led to by it. Formal monitoring of communication between the pilot and intervention students is not documented by the records, so a diffusion of procedures cannot be entirely excluded.

Implementation Fidelity and Practicality

All 150 expected final recordings were submitted (30 students × five cycles), and both the pre- and posttests were completed by all 30 students. Additional technical support during at least one cycle was required by four students. Overall participation was summarised as 95% by the original report, but separate numerators and denominators for attendance, peer-feedback completion, revision uptake, or lecturer-response timing were not provided by the retained manuscript. Only the directly supported completion evidence is therefore reported by the revision, and the unmeasured fidelity dimensions are explicitly identified.

Table 7. Fidelity evidence, interpretation, and limits

Fidelity dimension	Available evidence	Interpretation	Unresolved limitation
Recording completion	150/150 final cycle recordings submitted	Core production requirement was completed	Does not show number of drafts or replay frequency
Assessment completion	30/30 pretests and 30/30 posttests	Complete aggregate outcome dataset	Individual trajectories were not reported
Technical accessibility	Four students needed support; shared microphones/campus access used	Adaptation maintained access	Resource burden and time cost were not logged
Peer-feedback completion and quality	Template use was documented in class logs	Feedback procedure was implemented	Accuracy, specificity, and uptake were not systematically coded
Lecturer feedback timing	A 48-hour rule was specified and reported as implemented	Timeliness was an intended support	No numerator/denominator or timestamp audit

Fidelity dimension	Available evidence	Interpretation	Unresolved limitation
Moderation and privacy	Closed group, respectful-language rule, no public posting	Risk-reduction procedures were present	Rule violations and deletion schedule were not systematically audited

The total practicality scale was rated at $M = 4.18$ ($SD = 0.46$) by the students. The descriptive item-cluster means ranged from 4.10 for technological ease to 4.25 for guide clarity. The combined descriptive mean of the two lecturers was 4.38; inferential comparisons and lecturer standard deviations are not emphasised, because $n = 2$. Usability in the supported class context is suggested by these ratings, but acquiescence, social desirability, and researcher/instructor involvement may be reflected by them.

Table 8. Descriptive practicality results

Descriptive item cluster	Student M	Student SD	Lecturer combined mean ($n = 2$)
Guide clarity	4.25	0.52	4.50
Technological ease	4.10	0.61	4.25
Feedback usefulness	4.18	0.57	4.40
Learning and cultural relevance	4.20	0.54	4.35
Total descriptive scale	4.18	0.46	4.38

Preliminary Learning Change

The primary three-component speaking-performance score was increased from $M = 63.33$ ($SD = 8.10$) to $M = 81.00$ ($SD = 6.70$), with a mean difference of 17.67 points, $t(29) = 12.12$, $p < .001$, 95% CI [14.69, 20.65], $dz = 2.21$. Positive exploratory change was shown by each of fluency, pronunciation, and intonation. Observable speaking assurance was increased from 70.00 to 82.00, but it is not interpreted as psychological self-confidence. All dimension-level p values remained below the Holm-adjusted thresholds; nevertheless, the magnitude of change requires cautious interpretation, because a comparison group was lacked by the design, the raters were embedded in the course context, and performance may have been influenced by repeated exposure to the rubric.

Table 9. Pretest–posttest outcomes and construct interpretation

Outcome	Pre M (SD)	Post M (SD)	Mean difference	$t(29); p$	95% CI	dz
Speaking performance: 3 dimensions (primary)	63.33 (8.10)	81.00 (6.70)	17.67	12.12; $< .001$	[14.69, 20.65]	2.21
Fluency (exploratory)	60.00 (8.90)	78.00 (7.40)	18.00	10.20; $< .001$	[14.39, 21.61]	1.86
Pronunciation (exploratory)	62.00 (8.30)	80.00 (6.90)	18.00	11.50; $< .001$	[14.80, 21.20]	2.10
Intonation (exploratory)	68.00 (7.60)	85.00 (6.20)	17.00	9.80; $< .001$	[13.45, 20.55]	1.79
Observable speaking assurance (separate)	70.00 (8.00)	82.00 (6.80)	12.00	8.70; $< .001$	[9.18, 14.82]	1.59

Mixed-Methods Joint Display

Table 10. Integration of needs, design, implementation evidence, and quantitative interpretation

QUAL need / design inference	Instructional feature	Implementation evidence	Integrated inference
Every student needed more individual speaking time	Five required individual audio cycles	150/150 final recordings submitted	The sequence successfully distributed production opportunities; it does not establish that dosage alone caused gains
Students needed to hear and revise their own performance	Replay/self-review and opportunity to re-record	Self-review was required in the syntax; replay frequency was not logged	Self-monitoring is a plausible mechanism requiring direct process measurement
Feedback needed to be specific and usable	Peer prompts + lecturer rubric feedback	Template and response rule implemented; quality/uptake not coded	Interaction was structurally present, but feedback quality cannot explain effect magnitude
Technology quality was unequal	Low-bandwidth options and shared equipment	Four students received additional support	Practicality depended on institutional support, not only learner devices
Tasks needed meaningful content	Culturally responsive narrative and teaching tasks	High completion and positive relevance ratings	Cultural relevance was feasible, but no independent engagement effect was measured
Pretest–posttest change was large	Primary paired comparison	M increased 17.67 points; $d_z = 2.21$	Promising within-cohort change warrants controlled replication, not causal effectiveness claims

Discussion

It is suggested by the results that the feasibility of learner-generated podcasting is depended less on the audio medium itself than on the instructional architecture that surrounds it. Seven stages were linked by the developed model: the brief being interpreted and a topic being chosen, planning, recording, replaying and self-assessing, peer feedback being obtained, lecturer feedback being received, and revising or reflecting. It is indicated together by the expert review, the pilot experience, the practicality responses, the completion evidence, and the preliminary performance data that this sequence was workable in the target course. Feasibility was nevertheless conditional. The lowest practicality rating was received by technological ease, and microphones, appropriate recording spaces, or campus support were needed by some students. The large statistical improvement should therefore be read as evidence of promise within one cohort rather than as proof of causal effectiveness.

This interpretation is consistent with studies by which student podcasting is framed as authentic assessment rather than merely a different file format. It was reported by Wakefield et al. (2023) that an engaging assessment experience could be provided by learner-generated podcasts, while communication and collaboration were supported. Benefits for postgraduate science communication were found by Vainieri et al. (2024), but the continuing need for clearer task directions and production assistance was also noted. The value of podcasting as an alternative assessment was demonstrated by Halabi (2021), while its capacity to connect disciplinary learning, critical interpretation, and communication with audiences was emphasised by Moore (2024). In

foreign-language learning, it was shown by Acevedo de la Peña and Cassany (2024) that active production, multimodal composition, and engagement with authentic audiences can be stimulated by podcasting. It is added by the present study that educational value is generated through the combination of explicit criteria, repeated production, self-review, feedback, and compulsory revision.

The sequence also corresponds with research on formative assessment and digitally mediated feedback. It was shown by Cadena-Aguilar and Álvarez Ayure (2021) that awareness of comprehensibility could be heightened by the self- and peer-assessment of student podcasts. It was demonstrated by Xu and Peng (2022) that mobile-assisted peer feedback in speaking courses is shaped by motivation and participation quality. It is indicated by more recent studies that mobile peer feedback is most useful when comments are actively provided, interpreted, and applied by learners (Ding & Zhu, 2025; Weng et al., 2025). It was concluded by Kerman et al. (2024), however, that outcomes are depended on the interaction of technology, task design, learner characteristics, scaffolding, and teacher facilitation. For that reason, revision is a central stage in this model: limited instructional value is had by feedback unless it can be translated by students into an altered performance. This position is compatible with evidence that feedback literacy can be strengthened by self- and peer-assessment (Hoo et al., 2022) and that revision may be influenced in different ways by peer and teacher feedback (Cui et al., 2022).

In Indonesia, the findings are broadly compatible with Khotimah et al. (2022), by whom behavioural, cognitive, emotional, and agentic engagement in podcast-based speaking activities was described. Viewed through self-determination theory, engagement may be supported by podcast tasks when meaningful choice, competence development, and interaction are experienced by learners. An important condition is added by the current results: choice must be paired with structure and adequate technical support. It is likewise indicated by research on Indonesian online peer feedback that managed and scaffolded exchanges can be more productive than unsupported comments (Mulyati & Hadianto, 2023). Positive participation in a Southeast Asian online-feedback context was found by Waluyo and Panmei (2024), but engagement was not automatically associated with better learning. Such variation may be helped to be explained by differences in feedback quality, proficiency, task difficulty, teacher moderation, digital access, sensitivity to face-threatening criticism, and students' reliance on lecturers as authoritative evaluators.

The observed pretest–posttest change may be plausibly accounted for by three mechanisms. Speaking time may have been expanded and dependence on stronger group members may have been reduced by the requirement of individual recording. Pauses, pronunciation difficulties, grammatical problems, and fluency disruptions that are difficult to detect during real-time performance may have been helped to be noticed by students through replay against explicit criteria. Learners may also have been encouraged to convert evaluative information into specific adjustments by feedback that was followed by revision. These mechanisms were not directly measured. Maturation, concurrent coursework, task familiarity, rubric familiarity, instructor enthusiasm, Hawthorne effects, regression to the mean, or increasing comfort with recording technology could also have contributed. Explanatory propositions for future testing are therefore generated by the study, rather than mediation or causality being confirmed.

The culturally responsive component should be interpreted as a design condition, not as an independently verified intervention effect. Relevance may be increased by local experiences and narratives, but more than the insertion of local names or folklore is required by culturally responsive digital pedagogy. Attention to learner identity, digital participation, ownership of cultural resources,

representation, and control over communication is required by it (McDaniel, 2024). Multilingual resources are also recognised by culturally and linguistically responsive teaching as legitimate assets rather than deficits (Wesley-Nero & Donley, 2024). Learner agency, source accuracy, respectful representation, and reflection on audience and authorship are therefore preserved by the model. Because no validated engagement scale or cultural comparison condition was included, it cannot be concluded by the study that engagement was independently raised by culturally situated content.

An iterative interpretation of ADDIE is also supported by the findings. The product was returned to Design by the problems that were detected during expert appraisal and piloting, while further Development was informed by the implementation and evaluation evidence. Evaluation was functioned throughout the process rather than only at the end. From a self-determination perspective, it is suggested by the results that autonomy is more educationally productive when it is combined with competence-supporting criteria, accessible resources, and socially supportive feedback (Ryan & Deci, 2020). In practice, clear criteria, examples of constructive feedback, replay-based self-assessment, and a compulsory revision stage should be provided by lecturers. Audio production, source attribution, audience awareness, responsible AI use, and the ethical circulation of digital content should also be addressed by the guidance. Restricted sharing, private submission, flexible recording locations, and alternatives for students with anxiety or limited access are important safeguards.

At the institutional level, learner recordings should be treated as student data rather than as disposable class products. Tracking, unclear data practices, and limited control over collection and reuse may be of concern to students (Jones et al., 2020). Understandings of privacy and the willingness to adopt educational technologies can also vary across student groups (Kumi-Yeboah et al., 2023). Clear policies for consent, platform selection, ownership, storage duration, access, redistribution, and deletion should be established by institutions. Participation can be prevented from depending on personal economic resources by loanable recording devices, accessible facilities, and secure institutional platforms. Legitimate generative-AI support—such as limited transcription or pronunciation practice—should additionally be distinguished by the policies from the replacement of students' own oral performance.

The principal novelty of the study is the platform-neutral specification of learner-generated audio as a testable instructional sequence. The proposed seven-stage cycle appears feasible when four conditions are operated together: ethically restricted sharing, explicit performance criteria, scaffolded and moderated feedback, and accessible technical alternatives. Methodologically, needs evidence, expert appraisal, pilot findings, implementation records, practicality responses, and performance data were allowed by the qualitatively prioritised mixed-methods design to inform successive revisions. The contribution is therefore a theoretically grounded design proposition, by which the components to be retained, manipulated, and measured in future evaluations are identified.

Interpretation and transferability are restricted by several limitations. One convenience-selected class of 30 students from a single program was involved in the project, and no comparison group was included. A potential developer–evaluator conflict was created by the developer's involvement in evaluation. A formal saturation record, verified quotations, and a systematic negative-case register were lacked by the qualitative archive. Only three validators were included in the expert appraisal, and validator-specific scores, item-level agreement, uncertainty estimates, and documented blinding during re-appraisal were not retained. Course-based self-report was relied on

by the practicality evidence, and dimension-level reliability and full item distributions were unavailable. Content organisation, audience awareness, and validated affective outcomes were not fully covered by the speaking measure. Delayed retention, workload, replay frequency, feedback quality, revision uptake, external AI use, and data retention were not systematically examined. Multiple campuses, independent evaluators, randomised or quasi-experimental comparisons, delayed posttests, validated engagement and well-being measures, and detailed process evidence should be included in future work. The components that produce sustainable improvement would be helped to be identified by comparisons among conventional presentations, simple record-and-submit tasks, and the full seven-stage sequence.

Conclusion

A platform-neutral interactive podcast sequence for a university Speaking Skills course was developed and preliminarily evaluated by this study. The product is defined by a repeatable practice–replay–feedback–revision cycle, culturally responsive task criteria, and safeguards for privacy and access, rather than by a specific application. It was indicated by the expert appraisal and the user responses that the sequence was feasible in the supported class, and a large pretest–posttest improvement in a limited three-component speaking measure was demonstrated by the cohort. Because the evidence was obtained from a one-group developmental trial with incomplete process and psychometric data, the appropriate conclusion is feasibility and preliminary promise, not established effectiveness. In future evaluation, the seven-stage cycle, restricted sharing, criterion-referenced feedback, moderation, technical alternatives, and culturally responsive criteria should be preserved, while multiple classes, independent blinded raters, validated affective measures, complete fidelity indicators, individual change displays, and delayed follow-up are used. The proposed mechanisms would be tested rather than assumed by such a design.

References

- Acevedo de la Peña, I., & Cassany, D. (2024). Student podcasting for foreign language teaching–learning at university. *Journal of Technology and Science Education*, 14(1), 123–141. <https://doi.org/10.3926/jotse.2509>
- Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer. <https://doi.org/10.1007/978-0-387-09506-6>
- Cadena-Aguilar, A., & Álvarez Ayure, C. P. (2021). Self- and peer-assessment of student-generated podcasts to improve comprehensibility in undergraduate EFL students. *Profile: Issues in Teachers' Professional Development*, 23(2), 67–85. <https://doi.org/10.15446/profile.v23n2.88928>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Cui, Y., Schunn, C. D., & Gai, X. (2022). Peer feedback and teacher feedback: A comparative study of revision effectiveness in writing instruction for EFL learners. *Higher Education Research & Development*, 41(6), 1838–1854. <https://doi.org/10.1080/07294360.2021.1969541>
- Ding, Y., & Zhu, J. (2025). Mobile based peer feedback in EFL speaking: Learners' motivation, behavioral engagement in feedback provision, and achievement. *Asian-Pacific Journal of Second and Foreign Language Education*, 10, Article 10. <https://doi.org/10.1186/s40862-024-00314-9>
- Gay, G. (2018). *Culturally responsive teaching: Theory, research, and practice* (3rd ed.). Teachers College Press.
- Halabi, A. K. (2021). Pivoting authentic assessment to an accounting podcast during COVID-19. *Accounting Research Journal*, 34(2), 156–168. <https://doi.org/10.1108/ARJ-08-2020-0219>

- Hoo, H.-T., Deneen, C. C., & Boud, D. (2022). Developing student feedback literacy through self and peer assessment interventions. *Assessment & Evaluation in Higher Education*, 47(3), 444–457. <https://doi.org/10.1080/02602938.2021.1925871>
- Huang, H., & Li, M. (2024). The impact of technology-enhanced language learning environments on second language learners' willingness to communicate: A systematic review of empirical studies from 2012 to 2023. *Language Learning & Technology*, 28(1), 1–22. <https://doi.org/10.64152/10125/73593>
- Huang, H.-T. D. (2023). Examining the effect of digital storytelling on English speaking proficiency, willingness to communicate, and group cohesion. *TESOL Quarterly*, 57(1), 242–269. <https://doi.org/10.1002/tesq.3147>
- Jones, K. M. L., Asher, A., Goben, A., Perry, M. R., Salo, D., Briney, K. A., & Robertshaw, M. B. (2020). "We're being tracked at all times": Student perspectives of their privacy in relation to learning analytics in higher education. *Journal of the Association for Information Science and Technology*, 71(9), 1044–1059. <https://doi.org/10.1002/asi.24358>
- Kerman, N. T., Banihashem, S. K., Karami, M., Er, E., van Ginkel, S., & Noroozi, O. (2024). Online peer feedback in higher education: A synthesis of the literature. *Education and Information Technologies*, 29(1), 763–813. <https://doi.org/10.1007/s10639-023-12273-8>
- Khotimah, K., Cahyono, B. Y., & Batunan, D. A. (2022). Chronicling Indonesian EFL students' engagement in podcast-based speaking activities in online learning milieu: A self-determination theory perspective. *The JALT CALL Journal*, 18(3), 335–359. <https://doi.org/10.29140/jaltcall.v18n3.621>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kumi-Yeboah, A., Kim, Y., Yankson, B., Aikins, S., & Dadson, Y. A. (2023). Diverse students' perspectives on privacy and technology integration in higher education. *British Journal of Educational Technology*, 54(6), 1671–1692. <https://doi.org/10.1111/bjet.13386>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t* tests and ANOVAs. *Frontiers in Psychology*, 4, Article 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Little, T., Dawson, P., Boud, D., & Tai, J. (2024). Can students' feedback literacy be improved? A scoping review of interventions. *Assessment & Evaluation in Higher Education*, 49(1), 39–52. <https://doi.org/10.1080/02602938.2023.2177613>
- Man, D., Kong, B., & Chau, M. H. (2024). Developing student feedback literacy through peer review training. *RELC Journal*, 55(2), 408–421. <https://doi.org/10.1177/00336882221078380>
- McDaniel, D. S. (2024). Toward culturally digitized pedagogy: Informing theory, research, and practice. *Reading Research Quarterly*, 59(2), 193–210. <https://doi.org/10.1002/rrq.534>
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications.
- Moore, T. (2024). Pedagogy, podcasts, and politics: What role does podcasting have in planning education? *Journal of Planning Education and Research*, 44(3), 1134–1147. <https://doi.org/10.1177/0739456X221106327>
- Mulyati, Y., & Hadianito, D. (2023). Enhancing argumentative writing via online peer feedback-based essay: A quasi-experiment study. *International Journal of Instruction*, 16(2), 195–212. <https://doi.org/10.29333/iji.2023.16212a>
- Perry, S. (2024). Podcasting as a form of assessment: Increasing student motivation in academic English-speaking assessment. *Language Testing in Asia*, 14, Article 15. <https://doi.org/10.1186/s40468-024-00288-y>
- Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, 30(4), 459–467. <https://doi.org/10.1002/nur.20199>
- Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61, Article 101860. <https://doi.org/10.1016/j.cedpsych.2020.101860>

- Vainieri, I., Thackeray, L., Hillman, S., Perez, A., Roberts, R., & Panagiotopoulou, E. (2024). Evaluating podcasts as a science communication assessment for postgraduate students. *Innovations in Education and Teaching International*, 61(6), 1446–1459. <https://doi.org/10.1080/14703297.2023.2267047>
- Wakefield, A., Pike, R. K., & Amici-Dargan, S. (2023). Learner-generated podcasts: An authentic and enjoyable assessment for students working in pairs. *Assessment & Evaluation in Higher Education*, 48(7), 1025–1037. <https://doi.org/10.1080/02602938.2022.2152426>
- Waluyo, B., & Panmei, B. (2024). Students' peer feedback engagements in online English courses facilitated by a social network in Thailand. *Journal of Technology and Science Education*, 14(2), 306–323. <https://doi.org/10.3926/jotse.2305>
- Weng, J., Noordin, N., Ahmad, N. K., & Chen, L. (2025). A systematic review of mobile-assisted peer feedback in enhancing speaking proficiency for EFL learners. *Cogent Education*, 12(1), Article 2504228. <https://doi.org/10.1080/2331186X.2025.2504228>
- Wesley-Nero, S., & Donley, K. (2024). Culturally and linguistically responsive pedagogy: Examining teachers' conceptualizations of affirmative instructional practices for multilingual learners. *TESOL Journal*, 15(4), Article e881. <https://doi.org/10.1002/tesj.881>
- Xu, Q., & Peng, H. (2022). Exploring learner motivation and mobile-assisted peer feedback in a business English speaking course. *Journal of Computer Assisted Learning*, 38(4), 1033–1045. <https://doi.org/10.1111/jcal.12660>
- Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal*, 11(2), 49–54. <https://doi.org/10.21315/eimj2019.11.2.6>