AI-ASSISTED WEB-BASED CORRECTIVE FEEDBACK TO IMPROVE SENIOR HIGH SCHOOL STUDENTS' LISTENING COMPREHENSION

Marie Louise Catherine Widyananda[1]; Murti Kusuma Wirasti[2]; Dwi Kusumawardani[3]
[1,2,3]Master of Educational Technology, Faculty of Education, Universitas Negeri Jakarta, Indonesia
[1]Corresponding E-mail: cwidyananda@gmail.com

Abstract
Listening receives limited class time at the senior high school level, and large classes rarely allow teachers to give personal written feedback on students' open-ended answers. Web-based listening tools and AI scoring systems both exist, but the two are seldom integrated so that learners receive immediate corrective feedback on free-text listening responses. This study addressed that gap by developing and testing a web-based learning system assisted by artificial intelligence (AI), and it examined both the feasibility and the effectiveness of the product. The Research and Development method was applied through the Integrative Learning Design Framework across three phases, namely exploration, enactment, and evaluation. The product is a website named Ear Up! that delivers CEFR-graded audio lessons, collects essay answers, and uses an OpenAI GPT-4o model to return automated corrective feedback and next-lesson recommendations. The system was trialed with 28 tenth-grade students. Expert and user evaluation produced a mean feasibility score of 4.67 out of 5, in the very feasible band. Under a one-group pretest-posttest design, listening for details improved at a medium level (N-gain = 0.67) and listening for inference at a low level (N-gain = 0.25), while classical mastery rose from 29.41% to 82.35% and from 32.35% to 67.65%. The findings indicate that automated corrective feedback can extend listening practice beyond limited class time, yet higher-order inference still depends on teacher mediation. AI-assisted corrective feedback is therefore best positioned as a teacher-supervised supplement in future language learning.
Keywords: *Artificial Intelligence; Automated Corrective Feedback; Listening Comprehension; Web-Based Learning; Senior High School*

A. Introduction

Listening is the foundation of foreign language learning. Learners receive most of their language input through listening, and this input drives the acquisition of new vocabulary, grammar, and pronunciation (Gilakjani & Sabouri, 2016b; Alzamil, 2021). Listening comprehension is an active cognitive process in which a learner identifies sounds, retrieves prior knowledge, and constructs meaning from spoken text. When learners listen well, they also speak and write more confidently. For this reason, listening deserves a central place in English instruction at the senior high school level.

Listening is also difficult to master. Learners struggle with fast speech, unfamiliar accents, limited vocabulary, and complex topics (Gilakjani & Sabouri, 2016a; Al-Nafisah, 2019). Two sub-skills are especially demanding. *Listening for details* requires learners to capture specific facts such as names, numbers, and dates. *Listening for inference* requires learners to draw logical conclusions from information that is not stated directly. These higher-order sub-skills receive little practice when class time is short, and recent work calls for a rethinking of how listening is taught (Rahman et al., 2023).

An observation at a private senior high school in South Tangerang confirmed these problems. Listening was scheduled only once every two weeks, and the schedule was often reduced because teachers had to cover the main curriculum. A baseline listening test showed that the class reached mastery on only 46.57% of the items overall, and that the lowest scores appeared on the sections that measure listening for details and listening for inference. The limited class time also prevented teachers from giving personal written feedback on the open-ended answers that students produced. Students reported that they wanted to know whether their skills had improved and that they wanted transcripts so they could check their own mistakes.

Feedback is a decisive factor in learning. Effective feedback clarifies the expected standard, shows learners where they went wrong, and guides the next step (Nicol & Macfarlane-Dick, 2006; Dawson et al., 2019). Automatic feedback systems have been proposed to deliver this guidance quickly and at scale (Cavalcanti et al., 2021; del Gobbo et al., 2023). However, conventional feedback in large classes is slow, inconsistent, and shaped by the teacher's available time and judgment (van der Kleij, 2019). When feedback is delayed or vague, learners cannot use it to repair their understanding, and the value of practice declines.

Web-based learning offers one way to extend practice beyond the classroom. A website is flexible, self-paced, and accessible at any time, so it suits supplementary listening practice (Richards, 2015; Nurohmawati & Arini, 2023). Online and web-based learning has expanded rapidly and shows positive effects on access, engagement, and persistence across diverse settings (Castro & Tumibay, 2021; Farrell, 2020; Lakhal et al., 2021; Makruf et al., 2022; Pham et al., 2023; Yu & Jee, 2021). Earlier studies report that web-based and internet-based listening tasks improve learners' comprehension and metacognitive awareness (Pei & Suwanthep, 2019; Modaressi & Jalilzadeh, 2020). A website therefore allows learners to listen, answer questions, and review material outside school hours and at their own pace.

Artificial intelligence adds a second capability, namely automatic analysis of free-text answers. Large language models with natural language processing can read an essay answer and return immediate corrective feedback (Pikhart, 2020; Shadiev & Feng, 2023). Automated systems can already score essays and short answers and can flag incoherent responses (Hussein et al., 2019; Süzen et al., 2020; Urrutia & Araya, 2023). Generative models such as ChatGPT extend this capability to open-ended English writing (Fitria, 2023; French et al., 2023). Recent studies report measurable gains and high learner uptake from large language model corrective feedback, although this evidence comes almost entirely from writing rather than listening (Li et al., 2024; Polakova & Ivenz, 2024; Rahimi et al., 2024). Automated corrective feedback acts as a form of digital scaffolding that helps learners detect and repair errors without waiting for the teacher (Fan, 2023; Yunus, 2020). Recent systems show that AI can support authentic and ubiquitous language learning and can raise learner engagement (Dalu et al., 2023; Jia et al., 2022; Neo et al., 2022).

A clear gap nonetheless remains. Reviews of AI in education show that most systems still target prediction, analytics, and scoring rather than the instructional feedback that learners actually act on (X. Chen et al., 2020; L. Chen et al., 2020; Ouyang et al., 2023; Zhai et al., 2021; Zawacki-Richter et al., 2019). Where feedback is automated, it has been developed mainly for writing and short-answer grading, and its use for classroom listening remains limited (González-Calatayud, 2021; Haggag et al., 2018; Sánchez et al., 2022; Vo et al., 2022). In listening specifically, web-based tools deliver audio and closed quizzes but seldom evaluate the open-ended responses through which detail and inference are expressed, and almost none return immediate corrective feedback on those responses at the senior high school level. The present study addresses that gap directly by integrating CEFR-graded web-based listening tasks with AI-generated corrective feedback aimed at two demanding sub-skills, namely listening for details and listening for inference.

To address this gap, the study developed and evaluated the web-based AI listening system and was guided by three research questions. RQ1: How is a web-based AI system for listening comprehension developed through the exploration, enactment, and evaluation stages of the Integrative Learning Design Framework? RQ2: How feasible is the system according to expert validation and student perception? RQ3: How effective is the system in improving students' listening for details and listening for inference? The first question concerns the product, the second its quality, and the third its preliminary impact, so the three questions map directly onto the method and the findings that follow.

B. Method

Research design

This study used the Research and Development (R&D) method, which produces a product and tests its quality (Saragih et al., 2024). The development model was the Integrative Learning Design Framework (ILDF) proposed by Bannan-Ritland (2003). The original ILDF comprises four phases, namely informed exploration, enactment, evaluation of local impact, and evaluation of broad impact (Bannan-Ritland, 2003). This study implemented the first three phases, relabeled here as exploration, enactment, and evaluation, and did not carry out the fourth phase. The broad-impact phase concerns wide diffusion, multi-site adoption, and the generalization of the design theory, which lay beyond the scope of a single-school development project that was bounded by time and resources (Amalia et al., 2024). No phase was merged or dropped within the design cycle itself; rather, the study completed one full local design-and-evaluation loop and left broad-impact evaluation to future research. The exploration phase identified the real problem through a needs analysis, a survey of the literature, theory development, and audience characterization. The enactment phase turned the design into a concrete product through system design, an articulated prototype, and a detailed design of the website and its AI feedback engine. The evaluation phase judged the product through expert validation, a user trial, and an effectiveness test. Figure 1 shows the adapted model and its main steps.

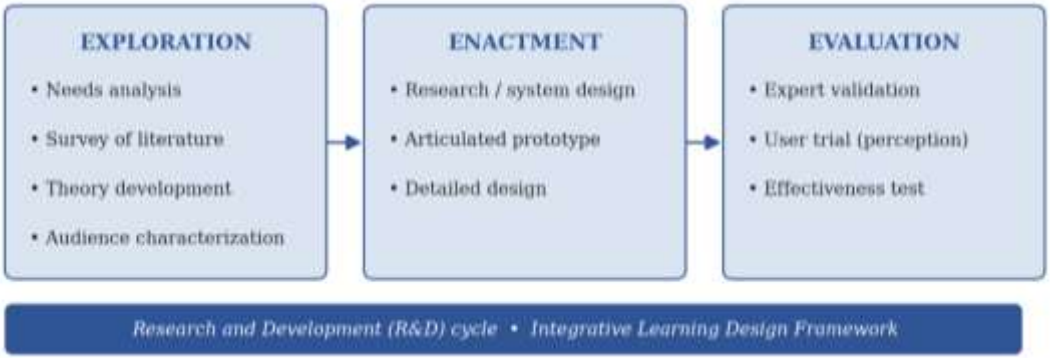

Figure 1 The adapted ILDF development model

Participants and setting

The study took place at a private senior high school in South Tangerang, Banten, from March 2024 to May 2025. The participants were 28 students in one grade-ten class, selected by purposive sampling because this class showed the listening difficulties found in the needs analysis. The students owned digital devices and had stable internet access, which is increasingly common among Indonesian students (Hidayat et al., 2022), so they were able to use the website at school and at home. In addition, four experts in instructional design, media, content, and language took part as validators. The learning content followed the listening competencies of the national Kurikulum Merdeka and targeted the two weakest sub-skills, namely listening for details and listening for inference.

Instruments

Data were collected with four instruments. First, an observation sheet recorded the listening activities and the difficulties that appeared in class, and it was built on listening-comprehension theory (Wilson, 2008; Goh & Vandergrift, 2022). Second, a questionnaire of eleven indicators measured the students' perception of the website on a five-point Likert scale, and it covered usability, content, and the AI feedback (Hidayat et al., 2022). Third, a performance test measured listening comprehension through open-ended questions. The test followed the B1 Preliminary for Schools format, was administered as a pre-test and a post-test, and measured the two target sub-skills with six items each. Each answer was scored from 0 to 3 on a rubric aligned with the CEFR descriptors. To check the validity of the automatic scoring, the class teacher independently re-scored a subset of the answers on the identical 0 to 3 rubric, and the teacher's scores were compared with the model's scores to confirm their agreement. Fourth, expert validation sheets collected judgments from the four specialists on the instructional design, media, content, and language of the product before the field trial.

Validity and reliability

The content validity of the test and the questionnaire was examined with the content validity ratio (Wilson et al., 2012). Items with a positive ratio were retained, and items with

a low ratio were revised or removed. The internal consistency of the questionnaire was estimated with Cronbach's alpha, and the questionnaire met the reliability criterion, with a coefficient above 0.70. The expert validation gave an additional content check. On the re-scored answers, the AI scores and the teacher scores were closely aligned, which supported the validity of the automatic assessment.

Data analysis

Data were analyzed with descriptive statistics. Feasibility was expressed as the mean of the expert and user ratings, which was then converted to a percentage and interpreted on a feasibility scale, where 81 to 100 percent means very feasible, 61 to 80 percent means feasible, and 41 to 60 percent means fairly feasible. Effectiveness was tested with a one-group pretest-posttest design and measured with the normalized gain (N-gain). The N-gain was computed as g = (post-test - pre-test) / (maximum score - pre-test) for each student and then averaged for the class (Hake, 1998). The value was read in three bands, namely high for a value of at least 0.7, medium for a value from 0.3 to below 0.7, and low for a value below 0.3. Classical mastery was reported as the percentage of students who reached the mastery threshold of 75 percent, and it was compared before and after the intervention. Because the design used a single group with no control group, the gain scores describe the change in this class over the intervention but cannot by themselves separate the effect of the system from other influences; the effectiveness results are therefore interpreted as preliminary evidence rather than as proof of a causal effect.

C. Finding and Discussion

Finding

The development produced a web-based AI listening system named Ear Up! The findings are reported by phase, namely the exploration phase, the enactment phase, and the evaluation phase.

Exploration phase

The needs analysis confirmed a clear gap in listening instruction. The baseline listening test followed the B1 Preliminary for Schools format and contained four sections. Figure 2 shows the percentage of students who reached mastery in each section. The class performed best on Part 1, which measures listening for details with short monologues, and performed worst on Part 3 and Part 4, which demand accurate detail capture and inference from longer texts. The overall mastery of 46.57% fell far below the threshold of 75%. The analysis also showed that limited class time prevented teachers from giving personal written feedback on the students' open-ended answers.

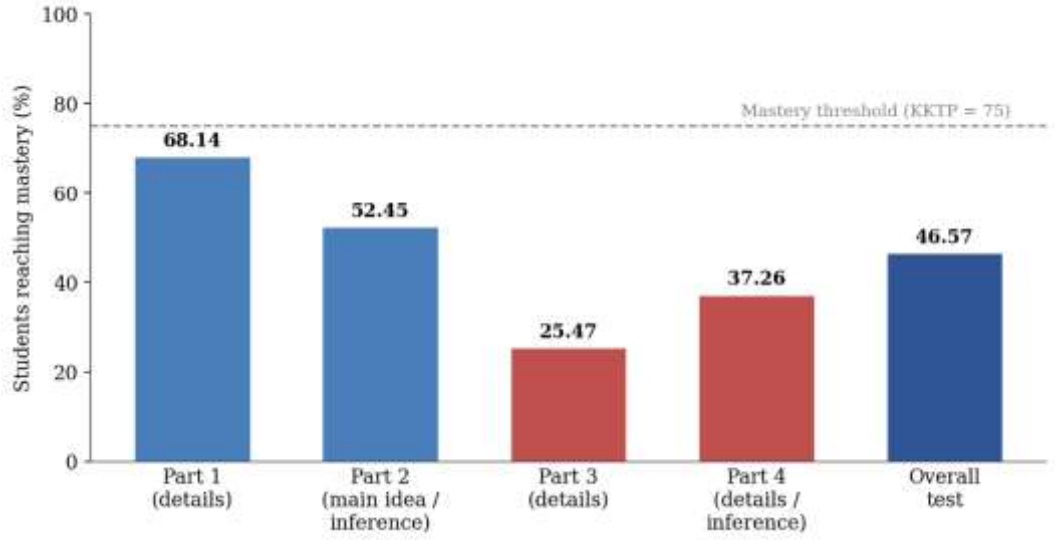

Figure 2 Students' baseline listening mastery by test section

The survey of the literature and the theory development pointed to a web-based solution that is integrated with AI. Web-based learning was chosen for its flexibility and its support for self-paced study (Richards, 2015; Pei & Suwanthep, 2019). The website was then integrated with a large language model so that it could analyze essay answers and return automated corrective feedback (Shadiev & Feng, 2023). The audience characterization described the students as digital natives who learn independently but who need instant confirmation of their progress.

Enactment phase

The learning design followed the achievement standard of the Kurikulum Merdeka and set two learning objectives. The first objective was to identify specific information such as names, locations, and key words, which corresponds to listening for details. The second objective was to draw a logical conclusion supported by at least two pieces of evidence, which corresponds to listening for inference. The learning flow adopted three listening stages, namely pre-listening, while-listening, and post-listening (Wilson, 2008), and used task-based listening lessons with authentic, CEFR-graded input (Goh & Vandergrift, 2022). Table 1 lists the topics, their CEFR levels, and their sources.

Table 1 Graded listening topics and sources in the Ear Up! system

No.	CEFR level	Topic	Source
1	A2	Favorite Foods	elllo.org
2	A2	Saving Money	elllo.org
3	B1	Allergy (placement)	Pathways 2A (Cengage)
4	B1	Indonesian Meals	elllo.org
5	B1	Cheating in Class	elllo.org
6	B2	Going Unplugged	elllo.org
7	B2	Overcoming Fear	elllo.org

No.	CEFR level	Topic	Source
8	C1	Standardized Test	elllo.org

The website assessed each open-ended answer with an OpenAI GPT-4o model. The model scored every answer against a four-criterion rubric on a scale from 0 to 3. Table 2 presents the rubric. The first three criteria measure listening comprehension, and the fourth criterion measures the clarity of the written response.

Table 2 AI scoring rubric for open-ended listening answers

Criterion	Score 3	Score 2	Score 1	Score 0
Identification of specific information	Captures all specific information accurately	Captures most information with minor gaps	Captures only a little information	Cannot identify specific information
Drawing conclusions	Draws a logical, well-supported conclusion	Draws a fairly logical conclusion	Draws a weak or partial conclusion	Cannot draw a conclusion
Use of supporting evidence	Cites strong, relevant evidence	Cites relevant but incomplete evidence	Cites weak or irrelevant evidence	Provides no supporting evidence
Clarity of expression	Expresses ideas clearly and coherently	Expresses ideas fairly clearly	Expresses ideas with poor organization	Provides no clear explanation

The AI feedback engine worked through prompt engineering. The model was given a clear role as a senior high school English teacher who addresses the student directly. It followed a fixed evaluation procedure, received a comprehensive instructional context, and worked under strict constraints. The context included the learning objectives, the scoring rubric, the audio transcript, the open-ended questions, and sample answers. The constraints set a minimum and maximum score, required a valid output format, and standardized the style and length of the feedback. Figure 3 shows the workflow that produces the feedback.

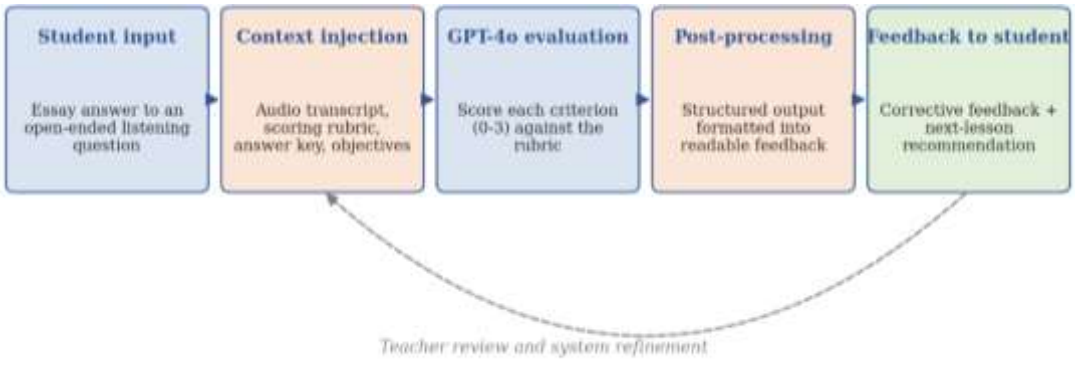

Figure 3 The web-based automated corrective feedback workflow

The website presented these functions through four design components. Table 3 summarizes the components and the features that students saw on screen. A separate teacher dashboard allowed teachers to monitor each student's answers, scores, and CEFR progress, which kept the teacher in control of the learning process.

Table 3 Blueprint of the web-learning components

Component	Description	Student-facing features
Instructional design	Structured pre-, while-, and post-listening activities with mastery learning	Competency map, learning flow, locked lessons, mastery markers
Multimedia	Audio stimulus and space for written answers	Audio player and a wide text area for essay answers
Internet technology	Web platform integrated with AI for automatic feedback	Submit button linked to the AI engine and an instant feedback box
Learning management	Individual tracking of progress and recommendations	Login, progress-tracking chart, and next-lesson recommendation panel

An illustrative example of the automated corrective feedback

To show how the system worked in practice, a representative listening-for-details exchange is reproduced below. The audio reported that a traveler missed his train because the roads were blocked by heavy traffic.

> *Typical student answer: "The man missed the train because he woke up late."*
> *Automated feedback (score 2 of 3): "Good effort. You correctly identified the main event, that the man missed his train. Look at the reason again, though. The speaker says the roads were blocked by heavy traffic, not that he overslept. Listen once more from the middle of the clip and notice the phrase about being stuck in traffic, then revise your answer to name the real cause."*

The exchange shows the immediate, specific, and standard-referenced character of the feedback. It confirms what the learner did well, locates the precise error against the transcript, directs attention to the relevant part of the audio, and sets a concrete next step. For inference items the same engine applies, but the model must judge a line of reasoning rather than a single fact, so its feedback is necessarily less precise, as the effectiveness results below confirm.

Evaluation phase

The feasibility evaluation drew on expert validation and on a user trial. The experts in instructional design, media, content, and language reviewed and validated the product before the field trial. The user trial collected the perception of the 28 students across eleven indicators, which is a common measure of online learning quality (Bayrak et al., 2020). Table 4 reports the mean scores. The overall mean was 4.67 out of 5, which falls in the very feasible

band. The students rated the systematic arrangement of the material and the authenticity of the audio most highly.

Table 4 Student perception of the Ear Up! system (n = 28)

No.	Perception indicator	Mean
1	Ease of using the website for listening practice	4.68
2	Smooth access and use across devices	4.68
3	Clarity of the language used in the material	4.64
4	Quality of the open-ended while-listening questions	4.71
5	Relevance of the topics to students' daily life	4.68
6	Authenticity of the conversations in the audio	4.79
7	Systematic arrangement of the listening stages	4.82
8	Ease of following the task instructions	4.46
9	Availability of instant AI feedback after submission	4.57
10	Clarity of the AI feedback on errors and gaps	4.68
11	Usefulness of the AI next-lesson recommendation	4.71
	Overall mean	4.67

The system was used by the class in two implementation rounds before the post-test. The effectiveness test then compared a pre-test and a post-test on the two target sub-skills. Each sub-skill was measured with six items, and the mastery threshold was a score of 4.5, which equals 75%. Table 5 and Figure 4 report the results. Listening for details improved from a mean of 3.6 to a mean of 5.2, which gives an N-gain of 0.67 in the medium band. Listening for inference improved from a mean of 4.0 to a mean of 4.5, which gives an N-gain of 0.25 in the low band. Classical mastery for listening for details rose from 29.41% to 82.35%, and classical mastery for listening for inference rose from 32.35% to 67.65%. The implementation also showed progress on the CEFR scale, because 5 of the 28 students (17.85%) advanced by one level, and one student reached the C1 level.

Table 5 Effectiveness of the system on the two target listening sub-skills (n = 28)

Sub-skill	Pre-test mean	Post-test mean	N-gain	Category	Mastery before	Mastery after
Listen for details	3.6	5.2	0.67	Medium	29.41%	82.35%
Listen and infer	4.0	4.5	0.25	Low	32.35%	67.65%

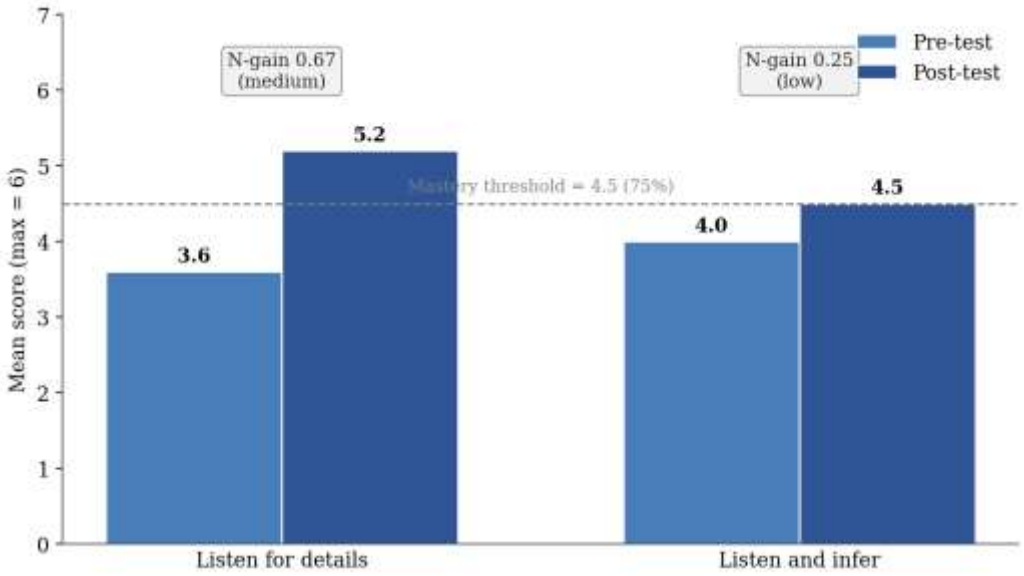

Figure 4 Pre-test and post-test means for the two target listening sub-skills

2. Discussion

The development followed the ILDF model and produced a working web-based AI system for listening. The three-phase adaptation kept the work focused on a clear local problem and on a concrete product (Bannan-Ritland, 2003; Saragih et al., 2024). The exploration phase grounded the design in real needs, the enactment phase built the product, and the evaluation phase tested its feasibility and effectiveness. This sequence shows that a design-based process can turn a classroom problem into a usable learning technology.

The high feasibility score supports the design choices. The mean user score of 4.67 indicates that students found the website easy to use and the material well organized. This result agrees with earlier reports that web-based listening tasks raise learner engagement and support self-paced study (Pei & Suwanthep, 2019; Nurohmawati & Arini, 2023). The students gave the highest ratings to the systematic listening stages and to the authentic audio, which confirms that a clear structure and real input help learners follow a complex task on their own.

The effectiveness results show that automated corrective feedback improved listening, but not equally across sub-skills. Listening for details improved by a medium gain, whereas listening for inference improved by only a low gain, and a few students even scored lower on the inference post-test. This pattern is the central limitation of the system and warrants closer analysis, because at least four factors plausibly converged. First, the task itself is harder, since inference requires learners to combine several pieces of spoken information and to reason beyond what is stated, so the room for short-term improvement is smaller than for factual recall (Gilakjani & Sabouri, 2016a; Rahman et al., 2023). Second, the feedback design favored detail, because for a factual item the model can verify an answer against the transcript and point to the missing fact, whereas for inference the target is a line

of reasoning rather than a fixed key, which is harder to evaluate and to correct in a single automated comment. Third, the intervention was short, spanning only two implementation rounds, and higher-order skills typically need sustained, distributed practice before they shift. Fourth, the students began from a low base in inference, so a brief automated comment could confuse as much as clarify when the underlying reasoning was still weak. The model's own limits compound these factors, because GPT-4o can misread implicature, idioms, and culture-specific cues and is therefore less reliable on inference than on concrete detail (Shadiev & Feng, 2023). Taken together, these factors indicate that automated feedback alone is insufficient for inference and must be paired with richer, teacher-guided scaffolding.

The role of feedback explains much of the improvement. Immediate and specific feedback clarifies the standard, shows the learner what to repair, and drives the next attempt (Nicol & Macfarlane-Dick, 2006; Dawson et al., 2019). The AI delivered this feedback at once, which conventional instruction could not do because of limited time (van der Kleij, 2019). In this sense the AI acted as a digital scaffold that extended the teacher's reach beyond the classroom (Fan, 2023; Yunus, 2020). The next-lesson recommendation also supported self-regulated learning by guiding students toward suitable material (Wong et al., 2019). Autonomous listening practice supported by AI has also been shown to raise learner confidence and effort (Hu & Hu, 2020).

The system reflects a wider movement to use artificial intelligence for language learning. Large language models can analyze free text and return natural, context-aware responses, which makes them useful for assessment and feedback (Pikhart, 2020; Jia et al., 2022). At the same time, AI feedback is not perfect. The model can misread idioms and culture-specific expressions, and its judgments are not absolute (Shadiev & Feng, 2023). Over-reliance on AI can also reduce student motivation when the design ignores the human side of learning (Sumakul & Hamied, 2023). The teacher dashboard in this study kept the teacher in the loop, which is the safer model in which AI complements rather than replaces the teacher.

These results carry a practical implication for how AI feedback should be positioned in the listening classroom. The most defensible model is a division of labor rather than a handover. The AI is well suited to high-frequency, low-stakes detail feedback, where it can give every student an immediate, rubric-referenced response that one teacher could not provide for a whole class within the limited lesson time. This frees teacher time for the work that the model does least well, namely guiding inference through questioning, modeling reasoning aloud, and discussing why a conclusion does or does not follow from the text. The teacher dashboard makes the balance workable in practice, because it lets the teacher review the automated feedback, correct it when the model misreads an implicature, and decide which students need face-to-face support. Positioned in this way, automated corrective feedback extends the teacher's reach without displacing the teacher's judgment, which is the balance that the present findings recommend.

The study has limitations. The trial involved one class of 28 students at a single school, so the results cannot be generalized to all learners. The intervention covered a short period, and the effectiveness test did not use a separate control group. The AI feedback was validated against a rubric and a teacher review, but it still requires human checking for high-stakes decisions. These limitations point to clear directions for future work.

D. Conclusion

This study contributes to English language teaching and educational technology a working model that integrates web-based, CEFR-graded listening practice with AI-generated corrective feedback on students' open-ended answers, a combination that earlier listening tools rarely offer. Built through the Integrative Learning Design Framework, the Ear Up! system proved feasible and improved the two target sub-skills to different degrees, with a clear gain for listening for details and a modest gain for listening for inference. The contribution lies less in any single score than in showing that immediate, rubric-based AI feedback can be embedded in routine listening practice and can extend a teacher's reach beyond the limited class time.

These findings come from one class of 28 students in a single school under a one-group design, so they should be read as preliminary rather than generalizable evidence. Within that limit, automated corrective feedback emerges as a promising supplementary tool for listening instruction. It is most dependable for factual, detail-level skills and least dependable for higher-order inference, which still requires teacher mediation. The system is therefore best deployed under teacher supervision, with the AI handling routine detail feedback and the teacher guiding inference. Future research should test the model with larger samples and a control group, design richer scaffolding for inference, and combine automated and teacher feedback, so that schools can adopt AI-assisted listening instruction in a balanced and pedagogically sound way.

E. Acknowledgment

The authors thank the school, the teachers, and the students who took part in this study. The authors received no specific funding for this research and declare no conflict of interest.

F. Bibliography

Al-Nafisah, K. I. (2019). Issues and strategies in improving listening comprehension in a classroom. *International Journal of Linguistics, 11*(3), 93. https://doi.org/10.5296/ijl.v11i3.14614

Alzamil, J. (2021). Listening skills: Important but difficult to learn. *Arab World English Journal, 12*(3), 366–374. https://doi.org/10.24093/awej/vol12no3.25

Amalia, D., Ariani, D., & Chaeruman, U. A. (2024). Pengembangan learning object "interaktivitas pembelajaran daring" berdasarkan the first principles of instruction

untuk mata kuliah designing e-learning. *Jurnal Pembelajaran Inovatif, 7*(1), 36–43. https://doi.org/10.21009/JPI.071.04

Bannan-Ritland, B. (2003). The role of design in research: The integrative learning design framework. *Educational Researcher, 32*(1), 21–24. https://doi.org/10.3102/0013189X032001021

Bayrak, D. F., Moanes, D. M., & Altun, D. A. (2020). Development of online course satisfaction scale. *Turkish Online Journal of Distance Education, 21*(4), 110–123. https://doi.org/10.17718/tojde.803378

Castro, M. D. B., & Tumibay, G. M. (2021). A literature review: Efficacy of online learning courses for higher education institutions using meta-analysis. *Education and Information Technologies, 26*(2), 1367–1385. https://doi.org/10.1007/s10639-019-10027-z

Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y.-S., Gašević, D., & Mello, R. F. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence, 2*, 100027. https://doi.org/10.1016/j.caeai.2021.100027

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access, 8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, 100002. https://doi.org/10.1016/j.caeai.2020.100002

Dalu, Z. C. A., Satrio, A., Aprastin, T. N. B., & Maulidah, S. (2023). Platform microlearning object berbantuan Open AI (artificial intelligence) sebagai upaya membangun lingkungan pembelajaran mandiri bagi mahasiswa pelaksana MBKM. *EPISTEMA, 4*(2), 154–165. https://doi.org/10.21831/ep.v4i2.66893

Dawson, P., Henderson, M., Mahoney, P., Phillips, M., Ryan, T., Boud, D., & Molloy, E. (2019). What makes for effective feedback: Staff and student perspectives. *Assessment & Evaluation in Higher Education, 44*(1), 25–36. https://doi.org/10.1080/02602938.2018.1467877

del Gobbo, E., Guarino, A., Cafarelli, B., Grilli, L., & Limone, P. (2023). Automatic evaluation of open-ended questions for online learning: A systematic mapping. *Studies in Educational Evaluation, 77*, 101258. https://doi.org/10.1016/j.stueduc.2023.101258

Fan, N. (2023). Exploring the effects of automated written corrective feedback on EFL students' writing quality: A mixed-methods study. *SAGE Open, 13*(2). https://doi.org/10.1177/21582440231181296

Farrell, O. (2020). A balancing act: A window into online student engagement experiences. *International Journal of Educational Technology in Higher Education, 17*(1). https://doi.org/10.1186/s41239-020-00199-x

Fitria, T. N. (2023). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. *ELT Forum: Journal of English Language Teaching, 12*(1), 44–58. https://doi.org/10.15294/elt.v12i1.64069

French, F., Levi, D., Maczo, C., Simonaityte, A., Triantafyllidis, S., & Varda, G. (2023). Creative use of OpenAI in education: Case studies from game development. *Multimodal Technologies and Interaction, 7*(8), 81. https://doi.org/10.3390/mti7080081

Gilakjani, A. P., & Sabouri, N. B. (2016a). Learners' listening comprehension difficulties in English language learning: A literature review. *English Language Teaching, 9*(6), 123. https://doi.org/10.5539/elt.v9n6p123

Gilakjani, A. P., & Sabouri, N. B. (2016b). The significance of listening comprehension in English language teaching. *Theory and Practice in Language Studies, 6*(8), 1670. https://doi.org/10.17507/tpls.0608.22

Goh, C. C. M., & Vandergrift, L. (2022). *Teaching and learning second language listening: Metacognition in action* (2nd ed.). Routledge.

González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. *Applied Sciences, 11*(12), 5467. https://doi.org/10.3390/app11125467

Haggag, M. H., Latif, M. A., & Helal, D. M. (2018). A learning analytics approach for student performance assessment. *International Journal of Computer Science and Information Technology, 10*(4), 79–94. https://doi.org/10.5121/ijcsit.2018.10407

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74. https://doi.org/10.1119/1.18809

Hidayat, D. N., Lee, J. Y., Mason, J., & Khaerudin, T. (2022). Digital technology supporting English learning among Indonesian university students. *Research and Practice in Technology Enhanced Learning, 17*(1). https://doi.org/10.1186/s41039-022-00198-8

Hu, J., & Hu, X. (2020). The effectiveness of autonomous listening study and pedagogical implications in the module of artificial intelligence. *Journal of Physics: Conference Series, 1684*(1), 012037. https://doi.org/10.1088/1742-6596/1684/1/012037

Hussein, M. A., Hassan, H., & Nassef, M. (2019). Automated language essay scoring systems: A literature review. *PeerJ Computer Science, 5*, e208. https://doi.org/10.7717/peerj-cs.208

Jia, F., Sun, D., Ma, Q., & Looi, C.-K. (2022). Developing an AI-based learning system for L2 learners' authentic and ubiquitous learning in English language. *Sustainability, 14*(23), 15527. https://doi.org/10.3390/su142315527

Lakhal, S., Khechine, H., & Mukamurera, J. (2021). Explaining persistence in online courses in higher education: A difference-in-differences analysis. *International Journal of Educational Technology in Higher Education, 18*(1), 1–32. https://doi.org/10.1186/s41239-021-00251-4

Li, J., Huang, J., Wu, W., & Whipple, P. B. (2024). Evaluating the role of ChatGPT in enhancing EFL writing assessments in classroom settings: A preliminary investigation. *Humanities and Social Sciences Communications, 11*(1), 1268. https://doi.org/10.1057/s41599-024-03755-2

Makruf, I., Rifa'i, A. A., & Triana, Y. (2022). Moodle-based online learning management in higher education. *International Journal of Instruction, 15*(1), 135–152. https://doi.org/10.29333/iji.2022.1518a

Modaressi, G., & Jalilzadeh, K. (2020). A comparative study of two ways of presentation of listening assessment: Moving towards internet-based assessment. *Language Teaching and Educational Research, 3*(2), 176–194. https://doi.org/10.35207/later.742121

Neo, M., Lee, C. P., Tan, H. Y.-J., Neo, T. K., Tan, Y. X., Mahendru, N., & Ismat, Z. (2022). Enhancing students' online learning experiences with artificial intelligence (AI): The MERLIN project. *International Journal of Technology, 13*(5), 1023–1034. https://doi.org/10.14716/ijtech.v13i5.5843

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. https://doi.org/10.1080/03075070600572090

Nurohmawati, M., & Arini, R. (2023). Designing web-based English listening tasks for university students. *Journal of Education and Teaching Learning (JETL), 5*(3), 51–62. https://doi.org/10.51178/jetl.v5i3.1529

Ouyang, F., Wu, M., Zheng, L., Zhang, L., & Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education, 20*(1), 4. https://doi.org/10.1186/s41239-022-00372-4

Pei, T., & Suwanthep, J. (2019). Effects of web-based metacognitive listening on Chinese university EFL learners' listening comprehension and metacognitive awareness. *Indonesian Journal of Applied Linguistics, 9*(2), 480–492. https://doi.org/10.17509/ijal.v9i2.20246

Pham, M., Singh, K., & Jahnke, I. (2023). Socio-technical-pedagogical usability of online courses for older adult learners. *Interactive Learning Environments, 31*(5), 2855–2871. https://doi.org/10.1080/10494820.2021.1912784

Pikhart, M. (2020). Intelligent information processing for language education: The use of artificial intelligence in language learning apps. *Procedia Computer Science, 176*, 1412–1419. https://doi.org/10.1016/j.procs.2020.09.151

Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education, 11*(1), 2410101. https://doi.org/10.1080/2331186X.2024.2410101

Rahimi, M., Fathi, J., & Zou, D. (2024). Exploring the impact of automated written corrective feedback on the academic writing skills of EFL learners: An activity theory perspective. *Education and Information Technologies, 30*(3), 2691–2735. https://doi.org/10.1007/s10639-024-12896-5

Rahman, M. N., Chowdhury, S., & Mazgutova, D. (2023). Rethinking teaching listening skills: A case study. *Language Teaching Research Quarterly, 33*, 176–190. https://doi.org/10.32038/ltrq.2023.33.10

Richards, J. C. (2015). The changing face of language learning: Learning beyond the classroom. *RELC Journal, 46*(1), 5–22. https://doi.org/10.1177/0033688214561621

Sánchez, C. J. P., Calle-Alonso, F., & Vega-Rodríguez, M. A. (2022). Learning analytics to predict students' performance: A case study of a neurodidactics-based collaborative learning platform. *Education and Information Technologies, 27*(9), 12913–12938. https://doi.org/10.1007/s10639-022-11128-y

Saragih, A., Mursid, R., & Sriadhi, S. (2024). Blended-based integrative learning design framework learning model: Improving numerical ability and competency in evaluation of student learning results. In *Proceedings of the 5th International Conference on Innovation in Education, Science, and Culture (ICIESC 2023)*. EAI. https://doi.org/10.4108/eai.24-10-2023.2342074

Shadiev, R., & Feng, Y. (2023). Using automated corrective feedback tools in language learning: A review study. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2022.2153145

Sumakul, D. T. Y. G., & Hamied, F. A. (2023). Amotivation in AI injected EFL classrooms: Implications for teachers. *Indonesian Journal of Applied Linguistics, 13*(1), 26–34. https://doi.org/10.17509/ijal.v13i1.58254

Süzen, N., Gorban, A. N., Levesley, J., & Mirkes, E. M. (2020). Automatic short answer grading and feedback using text mining methods. *Procedia Computer Science, 169*, 726–743. https://doi.org/10.1016/j.procs.2020.02.171

Urrutia, F., & Araya, R. (2023). Automatically detecting incoherent written math answers of fourth-graders. *Systems, 11*(7), 353. https://doi.org/10.3390/systems11070353

van der Kleij, F. M. (2019). Comparison of teacher and student perceptions of formative assessment feedback practices and association with individual student characteristics. *Teaching and Teacher Education, 85*, 175–189. https://doi.org/10.1016/j.tate.2019.06.010

Vo, N. N. Y., Vu, Q. T., Vu, N. H., Vu, T. A., Mach, B. D., & Xu, G. (2022). Domain-specific NLP system to support learning path and curriculum design at tech universities. *Computers and Education: Artificial Intelligence, 3*, 100042. https://doi.org/10.1016/j.caeai.2021.100042

Wilson, F. R., Pan, W., & Schumsky, D. A. (2012). Recalculation of the critical values for Lawshe's content validity ratio. *Measurement and Evaluation in Counseling and Development, 45*(3), 197–210. https://doi.org/10.1177/0748175612440286

Wilson, J. J. (2008). *How to teach listening*. Pearson Education.

Wong, J., Baars, M., Davis, D., Van der Zee, T., Houben, G.-J., & Paas, F. (2019). Supporting self-regulated learning in online learning environments and MOOCs: A systematic review. *International Journal of Human-Computer Interaction, 35*(4–5), 356–373. https://doi.org/10.1080/10447318.2018.1543084

Yu, J., & Jee, Y. (2021). Analysis of online classes in physical education during the COVID-19 pandemic. *Education Sciences, 11*(1), 3. https://doi.org/10.3390/educsci11010003

Yunus, W. N. M. W. M. (2020). Written corrective feedback in English compositions: Teachers' practices and students' expectations. *English Language Teaching Educational Journal, 3*(2), 95. https://doi.org/10.12928/eltej.v3i2.2255

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 1–27. https://doi.org/10.1186/s41239-019-0171-0

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J. B., Yuan, J., & Li, Y. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity, 2021*, 8812542. https://doi.org/10.1155/2021/8812542