



Castledown

 OPEN ACCESS

Technology in Language Teaching & Learning

ISSN 2652-1687

<https://www.castledown.com/journals/tltl/>

Technology in Language Teaching & Learning, 8, 104527 (2026)

<https://doi.org/10.29140/tltl.2026.104527>

Teacher-Managed Generative AI for Personalized Learning in Intermediate Japanese: A Classroom-Based Design Study of Reading Logs and Oral Information Sharing



TONG SUN 

Tokyo University of Foreign Studies, Japan
suntong95621@gmail.com

Abstract

This classroom-based design study examines how a teacher-managed, generative AI-supported learning system can be integrated into intermediate Japanese as a foreign language (JFL) instruction to support personalized learning and oral information sharing. While large language model-based generative AI can generate texts, explanations, questions, and feedback, its value in language teaching depends on how such affordances are aligned with pedagogical goals and classroom interaction. A teacher-managed web-based system was implemented over seven instructional sessions with five third-year Japanese majors. Data included system logs, task scores, timestamps, classroom observations, mini-interviews, and learner-produced presentation scripts. Findings from this exploratory design study suggest that personalization emerged through teacher-controlled prompt revision, learner-regulated pacing and sequencing, and immediate AI-generated feedback. Learning logs, interpreted together with classroom observations and mini-interviews, revealed iterative and expansive pathways of self-regulated learning. Presentation-script analysis showed that learners initially juxtaposed information from different texts, then stabilized information selection through same-text collaborative tasks, and later began to reorganize information under emerging higher-order categories. The study argues that generative AI is pedagogically meaningful not as an autonomous content generator, but as part of a teacher-designed learning ecology that connects individualized preparation, log-based formative assessment, and collaborative classroom use.

Keywords: generative AI, Japanese as a foreign language, personalized learning, learning logs, reading-to-speaking, computer-assisted language learning

Copyright: © 2026 Tong Sun. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. **Data Availability Statement:** All relevant data are within this paper.

Introduction

Intermediate learners of Japanese as a foreign language (JFL) often differ considerably in vocabulary knowledge, kanji reading ability, learning pace, topic interests, confidence in speaking, and experience with academic or semi-academic oral tasks. At this stage, instruction must move beyond the acquisition of discrete grammatical knowledge and support learners in using Japanese to understand, organize, and communicate information in socially meaningful ways. However, fixed classroom materials and uniform whole-class instruction do not always respond sufficiently to such learner diversity. Teachers are expected to select or develop materials that match learners' goals and levels, but the practical burden of continuously adapting texts, practice tasks, feedback, and classroom activities can be substantial.

Computer-assisted language learning (CALL) has long been introduced as one way to address learner diversity. CALL materials can offer multimodal input, repeated practice, self-paced learning, and immediate feedback (Chapelle, 2001; Warschauer, 1996). Yet many conventional CALL materials rely on predesigned content and predetermined feedback patterns. They can support repetition and individual pacing, but they are less able to generate new learning content dynamically, respond to unexpected learner needs, or provide feedback that reflects the specific context of a learner's response. These limitations are especially visible when teachers seek to connect individualized reading support with classroom-based oral communication.

Recent advances in large language model-based generative artificial intelligence (LLM-GAI) offer new possibilities for language education. Generative AI can produce texts, questions, explanations, summaries, and feedback in response to prompts, a flexibility that may help teachers respond to differences in learner interests, proficiency levels, and immediate difficulties. At the same time, AI-generated materials are not automatically pedagogically appropriate. Their educational value depends on prompt design, teacher evaluation, quality control, and integration into coherent instructional sequences. Without such design, generative AI may produce plausible but poorly aligned content, and learner-AI interaction may remain invisible to teachers.

Empirical research on generative AI in language education has expanded quickly, but the field remains uneven. Recent studies have examined ChatGPT and other forms of generative AI as support for second language writing, feedback, learner autonomy, teacher use in classroom preparation and implementation, and informal digital learning (Barrot, 2023; Choi et al., 2025; Liu & Ma, 2024; Strobl et al., 2024). Reviews have reported rapid growth in empirical work, while also noting that research remains concentrated in English language education, higher education, and writing-related skills (Li et al., 2025; Wang et al., 2025). Less attention has been paid to how generative AI can support non-English language learning, reading-based oral information sharing, and classroom designs that combine individual AI-supported preparation with collaborative classroom activity.

The present study addresses these gaps by developing and implementing a generative AI-supported Japanese learning system for intermediate JFL learners. The system was not used as a general-purpose chatbot or a stand-alone self-study tool. It was designed as a teacher-managed web environment with reading and speaking modules, prompt-controlled learning components, immediate feedback, and learning-log storage. The system was embedded in a backward-designed instructional sequence in which learners first engaged in individualized AI-supported reading activities and then used the information in pair or group presentation tasks.

The study is guided by two research questions. RQ1 asks: How did the generative AI-supported Japanese learning system enable personalization of learning content, pace, and feedback for intermediate JFL learners? RQ2 asks: How did learners' self-regulated learning processes and oral

information-sharing ability develop when AI-supported individual learning was integrated with collaborative classroom activity? By addressing these questions, the study shows how generative AI can be incorporated into a teacher-designed learning ecology that makes learner processes visible and connects personalized preparation with social language use.

Literature Review and Conceptual Framework

Generative AI in Language Education

Generative AI differs from earlier forms of CALL in that it can produce new language content and feedback in response to user input, which makes it attractive for language learning, where learners require varied input, meaningful output opportunities, and feedback responsive to specific attempts. Stockwell (2024), for example, argues that tools such as ChatGPT can assist teachers in generating topics, exercises, scenarios, and classroom activity ideas. In second language writing, generative AI has been examined as a writing buddy or revision support tool, with potential benefits for idea generation and language correction as well as risks related to overreliance and accuracy (Barrot, 2023; Strobl et al., 2024).

However, the current literature also raises concerns that are directly relevant to this study. First, many studies focus on general-purpose AI tools and learner perceptions rather than on how learner-AI interaction is monitored and integrated into instruction. Recent work has begun to foreground learner voices, showing that ChatGPT use in language education cannot be assumed and may be shaped by beliefs about tool limitations, ethical concerns, and broader views of language study (Hellmich et al., 2024). However, less is known about how learner-AI interaction unfolds when AI support is embedded in a teacher-managed classroom design and tracked through learning logs. Second, AI output quality is not guaranteed. Hallucination remains a structural problem of natural language generation, and educational use requires human verification, especially when texts contain factual information, dates, institutional details, or culturally specific knowledge (Ji et al., 2023). Third, generative AI research needs stronger methodological transparency, including descriptions of models, prompts, data sources, and procedures for evaluating AI-generated materials.

Personalized Learning, Self-Regulation, and Classroom Interaction

In this study, personalized learning refers to the flexible adjustment of learning content, learning pace, and feedback in relation to learners' interests, proficiency, progress, and difficulties. This definition avoids treating personalization as a purely technological function. Instead, personalization is understood as a pedagogical process mediated by technology but guided by teacher judgment. The AI system can expand the range and speed of possible adaptation, but the teacher remains responsible for defining learning outcomes, evaluating generated content, interpreting learner needs, and designing classroom activities.

Self-regulated learning provides a lens for analyzing how learners used the system. Self-regulated learning involves learners' active management of planning, monitoring, strategy use, evaluation, and adjustment (Pintrich, 2000; Zimmerman, 2002). In the present study, self-regulation was examined not as a stable trait but as an observable process in system logs. Repeated movement from practice tasks back to vocabulary learning, for example, can indicate monitoring and repair. Movement from foundational tasks toward optional speaking practice can indicate increased efficiency and confidence.

Classroom interaction is also central. Sociocultural perspectives emphasize that learning is mediated through tools, interaction, and participation in social practices (Lantolf & Thorne, 2006; Vygotsky, 1978).

In applying this perspective to an AI-supported classroom, however, human-human interaction must be distinguished from human-computer interaction. In this study, sociocultural principles were not operationalized by treating the AI system as a social partner or substitute for human interlocutors. Rather, the AI system was positioned as a mediating tool, a cultural artifact in the Vygotskian sense, whose primary role was to prepare and enable subsequent human-human interaction. Because the system generated reading texts on different topics for different learners, each learner possessed content that peers had not read. These AI-generated information gaps gave learners authentic reasons to question, explain, and negotiate meaning in the pair and group presentation tasks. Learner-AI interaction in the individual phase was therefore understood as tool-mediated preparation, while socially mediated knowledge construction was realized among learners in the collaborative phase. Sociocultural theory was thus operationalized at the level of the classroom design, in which AI-mediated individual preparation served as a precondition for peer interaction, rather than at the level of the human-computer dyad. AI-supported individual learning may help learners prepare, but oral information sharing requires learners to transform input into structured output for an audience. This transformation is supported by scaffolding (Wood et al., 1976), output opportunities (Swain, 1985), and staged task design that manages cognitive load (Sweller, 1988).

Backward Design and Reading-to-Speaking Development

Backward design was adopted to align learning goals, evidence of learning, AI-supported activities, and classroom tasks (Wiggins & McTighe, 2005). Rather than beginning with the technology, the teacher-researcher first identified target outcomes: consolidation of linguistic knowledge and development of oral information-sharing ability. Oral information-sharing ability was defined as the ability to select relevant information from reading texts, organize it coherently, and explain it clearly to peers. This definition suited learners with limited presentation experience, not yet ready for advanced argumentation or debate.

Japanese text and discourse studies informed the analysis of presentation scripts. Sakuma (1999) describes the discourse functions of the opening, development, and closing parts of Japanese texts, and Ishiguro (2020) emphasizes paragraph organization as central to textual clarity. These perspectives were used to examine whether learners merely extracted and juxtaposed source-text information or reorganized it under a coherent presentational structure. In this sense, the study connects international discussions of generative AI and CALL with Japanese-language-specific concerns about text structure, readability, and information organization.

Method

Research Design

This study employed an exploratory classroom-based design research approach. It was not an effectiveness study designed to prove that generative AI independently improved learners' language ability; rather, the purpose was to examine how a teacher-managed generative AI system, prompt-controlled learning materials, learning logs, and collaborative classroom tasks could be designed, implemented, and refined as a pedagogical model in an authentic instructional setting. The design was exploratory and formative: evidence from each session was used to adjust subsequent materials and activities, and the analysis focused on how the design supported personalization, process visibility, and reading-to-speaking development.

Multiple data sources were collected and triangulated: system logs, learner responses, timestamps, practice-task scores, classroom observations, mini-interviews, and learner-produced presentation

scripts. Logs provided behavioral traces of system use, interviews offered learner explanations, observations contextualized classroom interaction, and scripts evidenced oral information-sharing development.

Classroom observations were conducted through video recordings made with two fixed cameras at the front and back of the classroom. Their purpose was not to evaluate learners visually, but to contextualize the system logs and to check learning behaviors that could not be determined from the logs alone, such as note-taking, note consultation before reattempting a task, hesitation before moving on, and questions to the teacher or peers. These records were compared with log timestamps and mini-interview transcripts when interpreting learner pathways.

The final data corpus comprised 35 log files, of which 32 were generated during classroom sessions and three during learners' after-class self-study. The system also recorded six conversation-practice logs. Learner outcome data included 14 presentation scripts. In addition, seven group mini-interviews were conducted after the sessions; all five learners participated together in each interview, and the total interview time was approximately 140 minutes. The mini-interviews were audio-recorded and transcribed.

Context and Participants

The study was conducted at a university with five third-year undergraduate students majoring in Japanese. At the time of the implementation, the students had completed approximately 800 hours of university-level Japanese instruction and were studying upper-intermediate content in their regular curriculum. Their program included courses such as integrated Japanese, Japanese reading, Japanese literature, speech and debate, and business Japanese. Regular classes were generally conducted with approximately 20 students, which made it difficult for instructors to respond fully to individual differences in proficiency, pace, and communicative needs. Participants were recruited through open, voluntary enrollment. With the approval of the head of the Japanese department, a recruitment notice describing the purpose, content, and schedule of the implementation was distributed to all third-year students majoring in Japanese at the university; no students were individually approached or pre-selected. Students applied on the basis of their own interest and the compatibility of the session schedule with their timetables. Five students volunteered, and all five were accepted as participants; no further selection or screening criteria were applied. The participants therefore constitute a self-selected volunteer sample; this is taken into account when considering the transferability of the findings (see Limitations and Future Research).

Before the instructional implementation, the participants completed the SPOT 90 test as an approximate measure of Japanese proficiency. SPOT (Simple Performance-Oriented Test) was developed as a short method for estimating Japanese proficiency through rapid sentence-level processing (Kobayashi et al., 1996). The online SPOT 90 used in the Tsukuba Test-Battery of Japanese consists of 90 items and is intended to estimate practical Japanese processing ability under time pressure; the score guide interprets 31–55 as beginner, 56–80 as intermediate, and 81–90 as advanced (University of Tsukuba, n.d.). The participants' scores therefore indicated variation from upper-beginner to upper-intermediate levels, confirming the need for differentiated support. Participants are referred to by pseudonyms. Institutional information is withheld to protect anonymity. Participant information is summarized in Table 1.

The Generative AI-Supported Japanese Learning System

The learning system was a web-based Japanese learning environment developed by the teacher-researcher, using GPT-4o for text generation and feedback and speech recognition and synthesis models for the speaking module. Pedagogically, its defining feature was its two coordinated layers:

Table 1. *Participant Information*

Participant	Academic year	SPOT 90 score	Approximate proficiency level
A	Third year	53	Upper beginner
B	Third year	57	Intermediate
C	Third year	59	Intermediate
D	Third year	76	Upper intermediate
E	Third year	67	Intermediate

a learner-facing layer with a reading module and a conversation-practice module, and a teacher-facing management layer through which the teacher controlled what the AI generated, reviewed learner records, and revised content between sessions. This management layer was a central reason for building a dedicated system rather than having learners use general-purpose AI tools directly: it placed content generation, quality control, and process monitoring under instructional management.

Figures 1 and 2 show this learner-facing layer: on logging in, learners chose between the conversation-practice module and the reading-based Vocabulary and Grammar module used in this study (Figure 1), then selected a reading lesson from a topic menu (Figure 2).

A core function of the management layer was externalized content-generation control. Reading texts and the prompts controlling AI behavior were stored as editable files separate from the system itself, so the teacher could revise the learning components provided in subsequent lessons without rebuilding the system. For example, after the first session indicated that grammar practice was less necessary than vocabulary-reading practice and text organization, the teacher replaced grammar practice with vocabulary-reading practice and added a text-organization component.

Figure 3 shows a vocabulary reading-aloud item in progress, with the reading passage and questions at left and the learner's answers and the immediate feedback panel at right. The grammar-pattern learning tab (文型学習) remained available throughout; it was the separate grammar practice exercises, not grammar learning itself, that were removed after the first session.



Figure 1 *The Home Screen. Learners chose between The Conversation-Practice Module (会話練習) and the Reading-Based Vocabulary and Grammar Module (語彙と文法) Used in this Study. The Interface is in Japanese; English Glosses are Added for Presentation*

The system also generated immediate feedback. In reading-comprehension practice, rather than judging responses by keyword matching, it processed the learner's answer, the reading text, and prompt-based evaluation criteria together, enabling feedback that explained why an answer was inappropriate or how to improve it. When learners clicked the completion button, the system automatically saved the interaction history, learner responses, AI-generated feedback, and timestamps in log files organized by lesson topic. These logs formed a time-ordered portfolio of each learner's system use, which the teacher could review through the management layer when planning subsequent sessions.

Figure 4 sets out the full item-by-item feedback for this quiz, including the system's explanations of the four incorrect readings.

AI feedback on a vocabulary reading-aloud quiz
 花火大会 (Fireworks Festival) lesson — the activity shown in Figure 3. Each item asks for the correct reading of a word.
Score: 11 / 15 correct submitted answers: B D A C A C A D B B A B A C A · incorrect on items 2, 6, 8, and 14

#	Vocabulary	Reading	Gloss (EN)	Answer	Result
1	行事	gyōji	event	B	✓ correct
2	花火大会	hanabitaikai	fireworks festival	D	✗ incorrect
3	伝統	dentō	tradition	A	✓ correct
4	全国	zenkoku	nationwide	C	✓ correct
5	行われる	okonawareru	takes place	A	✓ correct
6	夜空	yozora	night sky	C	✗ incorrect
7	見どころ	midokoro	highlight	A	✓ correct
8	海辺	umibe	seaside	D	✗ incorrect
9	開かれる	hirakareru	is held	B	✓ correct
10	浴衣	yukata	yukata	B	✓ correct
11	会場	kaijō	venue	A	✓ correct
12	雰囲気	fun'iki	atmosphere	B	✓ correct
13	歓声	kansei	cheers	A	✓ correct
14	屋台	yatai	food stall	C	✗ incorrect
15	思い出	omoide	memory	A	✓ correct

The system's explanations for the incorrect items
 2. 花火大会 (hanabitaikai) — correct answer A: 「大会」 is read taikai, not daikai.
 6. 夜空 (yozora) — correct answer B: 「夜」 is read yo here, so 夜空 is yozora.
 8. 海辺 (umibe) — correct answer B: 「海」 is read umi here, so 海辺 is umibe.
 14. 屋台 (yatai) — correct answer A: a small shop at a festival; 「屋」 is read ya here.

Figure 4 A Readable Reconstruction of the System's Complete Feedback for the Activity in Figure 3, based on The Saved Learning-Log Data and with English Glosses. For Each of the 15 Vocabulary Items it Reports whether the Learner's Answer was Correct and, for Incorrect Items, gives the Correct Reading and a Short Explanation

Prompt Design and Materials Quality Control

Prompt design followed OpenAI's prompt-engineering guidance on clarity, specificity, and the structural separation of instructions from input data (OpenAI, n.d.). In practice, prompts specified the target level, topic, text length, expected structure, output format, question type, number of items, and feedback procedure. For reading-text generation, the prompt asked the model to create a Japanese reading passage on a specified topic for JLPT N3/CEFR B1 learners, approximately 480–520 Japanese characters in length, organized into an opening, development, and closing structure.

AI-generated reading texts were not used without teacher review. The appropriateness of AI-generated materials was evaluated from three perspectives: text difficulty, text structure, and information richness. Text difficulty was checked using the Japanese text readability measurement system, jReadability (Lee, 2016). Text structure was evaluated with reference to the three-part discourse structure of opening, development, and closing. Information richness was evaluated in terms of topical coverage, relevance, multiple perspectives, concreteness, depth of explanation, and factual accuracy. Proper nouns, dates, statistics, and institutional information were checked against reliable sources and revised when necessary.

Table 2 shows representative readability results from AI-generated reading texts evaluated during development. The scores ranged from 2.55 to 3.17 ($M = 2.80$), indicating that the generated texts generally fell within the intermediate readability range targeted in the study. These scores were not treated as proof of pedagogical appropriateness but as one source of evidence in the teacher's review and revision process.

Table 2. *Readability Scores of Representative AI-generated Reading Texts*

Topic	Characters	Readability score
Fireworks festival	527	2.80
Credit cards	512	2.82
Barcode payment	533	2.92
Shibuya Sky	499	2.62
University festivals	481	2.89
Ghibli Museum	534	2.55
Japanese-style pub	523	3.17
Anime & Manga	531	2.64

This quality-control procedure was necessary because LLMs can produce fluent but inaccurate or pedagogically misaligned output. In this study, prompt design provided an initial direction for generation, but teacher evaluation and revision were treated as an inseparable part of AI-supported materials development. Appendix A presents representative prompt excerpts.

Instructional Procedure

The implementation took place over seven sessions in the fall semester of 2025. The instructional model consisted of three stages: before class, during class, and after class. Before class, the teacher generated reading materials using AI, reviewed their appropriateness, revised them when necessary, and uploaded the approved materials to the server. This preparation reflected backward design because materials were selected and revised in relation to intended learning outcomes and evidence of learning.

During class, learners first selected a reading topic according to their interests and completed AI-supported individual learning. The teacher monitored learning logs and responded directly if learners had questions about AI feedback. After the individual phase, learners participated in peer-learning and presentation activities, designed to compensate for a limitation of AI-supported individual learning: AI could personalize preparation but did not by itself create social participation or shared knowledge construction.

After class, the teacher reviewed learning logs, task scores, errors, and presentation scripts to identify learner progress and difficulties; these data informed the revision of subsequent prompts and classroom tasks. Learners could also use the system for review and optional conversation practice related to the classroom topics.

Learning Content and Task Revision

The first session included vocabulary learning, grammar-expression learning, grammar practice, content-check questions, and reading-comprehension practice. However, mini-interview and observation data indicated that grammar practice was relatively easy because learners were accustomed to grammar memorization for university examinations and JLPT preparation. In contrast, learners reported difficulty reading vocabulary aloud and explaining reading content orally. In response, from the second session onward, grammar practice was removed, vocabulary-reading practice was added, and content-check questions were replaced by text-organization support.

Topics were also adjusted. At first, learners expressed broad interests such as Japanese culture and daily life in Japan but had difficulty specifying concrete topics. The teacher initially provided options based on situations Japanese university students might encounter. As the implementation progressed, learners began to articulate more specific interests, such as online payment methods, tourist sites discussed on social media, and the accuracy of online information about Japan. Later reading texts therefore covered a range of everyday and cultural topics, including defective product returns, train use, job hunting, apartment rental, fireworks festivals, credit cards, barcode payment, Shibuya Sky, university festivals, izakaya practices, two-dimensional culture, and the Ghibli Museum.

Data Analysis

For RQ1, content personalization was examined through changes in prompts, learning components, and reading topics; pace personalization through system logs showing time allocation, sequence, repetition, and movement among components; and feedback personalization through records of learner responses and AI-generated feedback.

For RQ2, system logs were reviewed as chronological records of learner actions. Learner actions were coded as initial study, practice completion, repetition, return to foundational learning after error or uncertainty, and movement toward advanced practice; operational definitions of these action codes are provided in Appendix B. Importantly, log entries were treated as behavioral indicators rather than direct evidence of learners' internal regulation. A navigation pattern was interpreted as monitoring, repair, or task-oriented adjustment only when supported by converging evidence, such as low or improved practice scores, timestamped repetition, classroom observation of note-taking or note consultation, or mini-interview comments explaining the reason for the change. Pathway classification relied on no quantitative thresholds but on a two-step qualitative procedure: the logs were first inspected for repetition or returns to foundational learning, and the reason for each pattern was then confirmed directly with the learner in the mini-interview before a label was assigned. Operational definitions and decision rules for the pathway categories and for repair are provided in the codebook in Appendix C. The coding was conducted by the author, which is appropriate for exploratory design research but is acknowledged as a limitation because independent coding was not conducted.

Presentation scripts were analyzed in relation to oral information-sharing ability. A formal codebook was developed for this analysis: every sentence in the AI-generated source texts and in each learner's presentation script was numbered and coded for its semantic function within its paragraph and within the text as a whole. The structural comparison described below was conducted on the basis of this

sentence-level coding, and the codebook with operational definitions is summarized in Appendix D. Each script was segmented into information units, defined as clauses or sentence groups that conveyed a distinct idea from the source texts. These units were compared with the source texts to identify retained information, omitted information, and newly reorganized information. The analysis then focused on two dimensions: information selection and organization, and logical coherence. Information selection and organization concerned whether learners selected important information from source texts and arranged it for listeners through grouping, sequencing, or categorization. Logical coherence concerned whether scripts showed a clear opening, development, and conclusion, and whether relationships among pieces of information were made explicit through topic sentences, connective expressions, and summary statements. All learner-produced scripts were originally written in Japanese; excerpts were translated into English by the author. Because of the small number of participants and the formative nature of the implementation, the analysis did not aim to establish causal effects or generalizable outcomes, but to develop a detailed account of how the AI-supported design functioned across learners, tasks, and sessions. Claims about learner development are therefore presented as design-based interpretations supported by converging evidence from logs, observations, interviews, scores, and learner-produced scripts.

Ethical Considerations

The relevant institutional authority confirmed in writing that the voluntary classroom-based instructional implementation did not require formal ethics committee review under the university's regulations, and departmental approval was obtained before recruitment. A recruitment notice was distributed to the relevant cohort, and only students who wished to participate joined the implementation. In this article, the term teacher-researcher refers to the author's dual role as designer of the learning system and instructor of the voluntary implementation sessions; it does not indicate an institutional teaching position at the participants' university. The author was not the instructor of the intermediate Japanese courses or of any other regular course in which the participants were enrolled, held no grading or supervisory authority over them, and had no institutional teacher-student relationship with them. The implementation was conducted outside the participants' regular curriculum, and participation and performance had no influence on their course grades, credits, or academic standing. All data collected during the implementation were accessible only to the author for research purposes and were kept confidential from the participants' regular course instructors and other university staff. During the orientation, the teacher-researcher explained the study purpose, procedures, data to be collected, and use of anonymized data for research. The data included system logs, learner-produced scripts, audio-recorded mini-interviews, and video-recorded classroom observations. All participants signed a written consent form before participating and were informed that participation was voluntary and that they could withdraw without penalty. Pseudonyms are used throughout this manuscript, and institutional information is withheld to protect anonymity. Consent forms and approval documents are not included as appendices to preserve participant privacy and the double-blind review process, but they can be provided confidentially to the editor upon request.

Findings

Personalization of Content, Pace, and Feedback

The first research question asked how the AI-supported system enabled personalization of content, pace, and feedback. The clearest evidence of content personalization was the revision of learning components after the first session — grammar practice was replaced by vocabulary-reading practice and text-organization support — made possible by the system's prompt-based design: changing the prompt files altered the learning content provided in later sessions without changing the base reading texts or interface.

Content personalization also occurred in topic selection. Learners initially lacked concrete topic preferences, but mini-interviews helped them articulate interests more specifically: after studying defective product returns, for example, they asked about online shopping and payment methods in Japan and about checking online information found on social media, and one learner later said, “I want to study using texts about Japanese anime.” These comments informed later reading topics such as credit cards, barcode payment, Shibuya Sky, and the Ghibli Museum.

Mini-interview comments also clarified why content personalization mattered to learners. One learner explained, “Vocabulary and grammar learning were almost the same as what we study in class, so I wanted to use AI to learn content that we cannot usually study in class.” Participants A and D, who both intended to study in Japan, similarly explained that the AI-generated materials felt tailored to their own needs and therefore more meaningful to study than the fixed content of textbooks, which strengthened their willingness to learn. Content personalization thus mattered not only through topical variety but through the perceived fit between materials and learners’ own goals.

Personalization of pace was visible in learners’ navigation patterns. The system did not require all learners to proceed through the same sequence at the same speed; personalization thus emerged not only from teacher adaptation but also from learner control over sequence, repetition, and time allocation. Participant D, who completed the required components efficiently and repeatedly moved on to free conversation practice, described this control: “When I study with the AI, I can lead my own learning at my own pace and in my own way. That feels really good, and for me it is more comfortable than learning in class.” The expansive pathway in this learner’s logs was thus matched by an explicit sense of ownership over pace and method. Learners also commented that AI-supported individual learning itself did not feel difficult; it felt less demanding than regular classroom learning because content they had not understood could be studied repeatedly, which is not possible in ordinary lessons. The difficulty they reported lay instead in the presentation tasks, consistent with the cognitive-load problems observed in the first information-gap session.

Feedback personalization was supported through immediate AI-generated responses. Because feedback was generated in relation to the learner’s answer, the reading text, and evaluation criteria, it could identify specific misunderstandings rather than simply reporting correct or incorrect answers. Feedback therefore functioned both as learner support and as evidence for instructional adjustment.

The effect of immediate feedback was most visible when learners received low scores and then chose to repeat or revisit earlier components. For example, Participant C scored 5/15 on vocabulary-reading practice in Session 2, repeated the same component, and improved to 14/15 before moving on to text-organization learning. Participant A similarly returned to vocabulary learning after scoring 6/15 in Session 3 and then improved to 13/15. Participant E inserted additional vocabulary review after weaker performance in the early sessions. Participant C’s mini-interview account corroborated this pattern, linking feedback immediacy to motivation: “As soon as I finished the practice, I could see the feedback and evaluation right away. Getting that immediate response made me more motivated to keep studying.” These cases suggest that feedback functioned not only as evaluation; it also appeared to prompt learner decisions to repair understanding before proceeding. Learner trust in the AI feedback was not unconditional, however. In one classroom episode, Participant C judged that the AI-provided answer to a reading-comprehension multiple-choice item was wrong and raised it directly with the teacher during class. The teacher reviewed the saved log on the spot, confirmed that the AI’s answer was correct, and resolved the doubt through explanation. The episode indicates that learners evaluated AI output critically rather than accepting it automatically, and that the stored logs allowed questions about AI quality to be verified and addressed in real time.

Learner Pathways Revealed by Logs

The pathway categories should be understood as triangulated interpretations rather than as direct inferences from clicks alone (see Appendix C). On this basis, the log analysis identified two broad tendencies in learner pathways, with several intermediate forms. The first was an iterative or compensatory pathway, in which learners returned from practice tasks to foundational learning after encountering low scores or uncertainty. The second was an expansive pathway, in which learners completed foundational tasks efficiently and moved toward optional conversation practice. Between these tendencies were stable-default and partial-compensatory pathways, where learners mostly followed the system sequence but made local adjustments when necessary.

Table 3 summarizes how these pathway tendencies appeared across the five participants and shows that self-regulation did not always appear as visible route changes. For some learners, maintaining the default sequence was itself an adaptive choice that matched their proficiency and task needs; for others, self-regulation appeared as repeated review, local repair, or movement toward optional speaking practice after the reading module.

For example, Participant A's return from vocabulary-reading practice to vocabulary learning was interpreted as compensatory adjustment because it followed a low score and was accompanied in the classroom observation by note-taking and later note consultation. By contrast, Participants B and D mostly maintained the default sequence; their stable scores, shorter completion times, and later movement to optional speaking practice suggested that the default sequence was already adaptive for their learning needs rather than a lack of self-regulation.

A classroom observation of Participant B further illustrates that movement toward optional conversation practice did not occur automatically. After completing the reading module in an earlier session, Participant B briefly attempted the conversation-practice module but stopped after one attempt, indicating that it was too difficult to continue independently. Rather than interpreting this non-use as a lack of learner motivation or as a limitation of the AI tool alone, the teacher encouraged the learner to try again gradually and treated speaking practice as an extension of reading-based preparation. In the later sessions, Participant B began to enter the conversation-practice module after completing the reading tasks, and by the final stage the learner was able to use free conversation practice. This episode shows that the pedagogical value of AI-supported speaking practice did not lie simply in making a conversation

Table 3. *Summary of Learner Pathways across Participants*

Participant	Dominant pathway	Log-based evidence and interpretation
A	Iterative/ compensatory	Returned after uncertainty; Session 3 vocabulary reading 6/15 -> vocabulary review -> 13/15. Later note-taking suggests strategy development.
B	Stable default -> expansion	Kept the default sequence with generally high scores; Session 7 included free conversation practice after reading completion.
C	Repair + task-oriented reordering	Repeated after low vocabulary-reading scores (5/15 -> 14/15; 6/15 -> 15/15). Later moved text-organization learning to support pair work.
D	Expansive/high proficiency	Highest SPOT 90 score (76); completed tasks accurately and moved to free conversation practice in multiple sessions.
E	Partial compensatory	Mostly followed the default sequence but inserted local vocabulary review; later used free conversation practice.

tool available. It emerged through teacher-mediated scaffolding, encouragement, and timing, which helped the learner cross a psychological and cognitive threshold from individualized reading support to oral practice.

From Individual Reading to Collaborative Information Sharing

The second research question concerned how self-regulated AI-supported learning connected with oral information-sharing development. The first session revealed that individual reading comprehension did not automatically transfer to oral explanation. When learners who had studied different texts were placed in an information-gap task, the task created a real communicative need but was too demanding at this initial stage: learners struggled to identify core information, read key vocabulary aloud, and reorganize content for listeners, and some simply combined source-text sentences with minimal reorganization.

This difficulty led to two revisions: vocabulary-reading practice was added so that learners could prepare for oral delivery, and text-organization support was added to help them grasp the three-part structure of a text and the relationship between main ideas and supporting details. Classroom tasks were also staged more carefully: later sessions used same-text collaborative work to reduce cognitive load, and once learners became more familiar with presentation organization, the class returned to more demanding integration tasks with related but different texts.

Development of Presentation Scripts

Presentation-script analysis showed development across three stages. In the initial different-text integration stage, learners were able to extract information from multiple source texts, but they often connected it physically rather than reorganizing it under a new logical framework. Table 4 illustrates this tendency: in both the fireworks and credit-card scripts, related pieces of information were placed together without a transition or organizing category linking them.

Table 4. *Stage 1: Juxtaposition of Information from Different Texts*

Topic	Source information	Learner script excerpt (translated)	Analytic interpretation
Fireworks festival	Text A described yukata as traditional summer clothing worn at fireworks festivals. Text B explained paid and free viewing seats.	“Fireworks festivals are divided into paid seats and free seats. Paid seats require tickets, and sheets are prepared by the river or in a square. Free seats can of course be used by anyone. However, to secure a good place, people wait from daytime. In the evening, many people wear yukata, that is, traditional Japanese clothing, and gather at the venue.”	The learner retained information from both texts but placed seating and clothing information in the same paragraph without an explicit organizing category or transition.
Credit cards	Text A described card types such as bank, shopping-site, and airline cards. Text B described application methods and screening.	“Sumitomo Mitsui Card, Rakuten Card, and Aeon Card are well known. Cards are also issued by supermarkets, shopping sites, and airlines. Credit cards are mainly applied for online, at banks, or at shopping mall counters. Online screening is popular because the result is fast.”	The learner physically combined information about card types and application methods, but the paragraph lacked a meaningful internal structure.

Table 5. *Stage 2: Improved Information Selection with Remaining Organizational Issues*

Topic	Improvement	Remaining issue
Barcode payment	The script retained service types, payment procedures, charging methods, large-city availability, rural limitations, and the need to carry cash.	Some information relations remained unclear, such as placing “Tokyo” and “WeChat Pay” at the same level.
Shibuya Sky	The script retained ticket purchase, price differences, daytime views, sunset waiting, and nighttime views.	Practical information and scenic attraction information were not clearly classified.
University festivals	The script retained food stalls, stage events, club activities, comedy/live events, celebrities, and examples such as the May Festival and Waseda Festival.	Events were mostly listed in parallel, and some wording still reflected the source text rather than audience-oriented reorganization.

The second stage used same-text collaborative tasks to stabilize information selection and organization. Because learners worked from a shared source text, they could focus on extracting main information, arranging it for listeners, and building a clearer presentation structure. As Table 5 shows, learners retained more core information than in the first stage, generally moving from general description to concrete examples.

At the same time, the second stage still showed limitations. Learners often listed information rather than classifying it under higher-order categories, treating items at different levels as parallel and not clearly separating practical information from scenic attractions (Table 5). Thus, same-text collaboration supported stability but did not fully develop hierarchical information organization.

In the third stage, learners returned to different-text integration tasks. Compared with the initial stage, scripts showed stronger attempts to reorganize information. The Ghibli Museum case was particularly informative (Table 6). Learners incorporated information from two source texts and, importantly, moved reservation information near the beginning of the script, before exhibitions and experiences. This suggests an emerging organizing principle: information needed before visiting was distinguished from information about experiences inside the museum.

Table 6. *Stage 3: Emerging Higher-Order Integration in the Ghibli Museum Case*

Higher-order frame	Information from source texts	How it appeared in the learner script	Evaluation
Practical information before visiting	No same-day tickets; advance reservation; early purchase.	Reservation information was moved near the beginning and foregrounded as information to know before visiting.	Relatively clear.
Experiences inside the museum	Exhibitions; short film; rooftop robot soldier; cafe; shop; limited products.	Exhibitions, film viewing, shopping, and rest were presented as a sequence of museum experiences.	Integration visible but still list-like.
Rules and cautions	Photography prohibited.	Photography rules were included within the experience-related information.	Functionally retained.
Cultural and touristic appeal	World of Ghibli works; popularity among foreign tourists.	Popularity and attraction information followed practical and experiential details, giving the museum touristic meaning.	Evidence of reordering.

Nevertheless, the third stage also showed that higher-order integration remained developing: some categories were still implicit, and several experiential and touristic details were presented consecutively. The scripts therefore showed progress from physical combination toward functional reorganization, but not yet full hierarchical classification. This characterization of development as movement toward hierarchical organization is itself an analytic interpretation, based on the author's codebook-guided reading of sentence order and grouping (Appendix D); the staged labels describe tendencies rather than discrete developmental levels.

Discussion

Generative AI as Teacher-Controlled Personalization

The findings should be interpreted in relation to the study's exploratory design research purpose. Rather than claiming that generative AI alone improved Japanese proficiency or oral performance, the study illustrates how a teacher-managed AI-supported ecology can connect individualized preparation, log-based process visibility, and collaborative classroom use. Its main contribution is therefore a pedagogical design model rather than a causal demonstration of AI effectiveness.

The findings suggest that the educational value of the AI-supported system lay less in automatic generation than in teacher-controlled personalization. This interpretation also aligns with recent conceptual work cautioning against treating human-AI interaction as full collaboration, since AI does not share purpose, reciprocal understanding, or distributed accountability in the same way that human participants do (Stockwell & Wang, 2025). For this reason, the pedagogical value of generative AI in the present study lay not in autonomous operation, but in its role within a teacher-designed ecology of materials, feedback, and classroom activity. The system allowed the teacher to adjust prompts, task types, and topics in response to learner needs, while keeping learner-AI interaction visible through logs.

Generative AI can produce abundant language-learning materials, but abundance alone does not ensure educational quality. The materials-evaluation procedure described earlier offers one way to make AI-generated Japanese reading materials more pedagogically accountable.

Learning Logs and the Visibility of Self-Regulation

The study also shows how learning logs can make traces of self-regulated learning visible when they are interpreted together with classroom observation and learner explanations. In many AI-supported contexts, teachers may see only the final product, not the process by which learners used AI feedback or revised their understanding. Here, logs showed learners' sequences, repetitions, scores, timestamps, and feedback histories, while video observation and mini-interview transcripts helped interpret why learners returned to earlier components. These combined traces made it possible to distinguish the four pathways and to interpret reduced learning time as possible strategic efficiency rather than necessarily reduced engagement.

The log data also show that self-regulated learning was socially oriented. Learners' choices in the individual AI-supported phase were shaped by the expectation of later explaining content to peers, so vocabulary-reading and text-organization practice became meaningful as preparation for oral information sharing. The system therefore supported individual regulation within a broader social task structure.

Integrating AI-Supported Reading with Collaborative Speaking Tasks

A key finding is that AI-supported reading alone did not automatically produce oral information-sharing ability; learners needed staged classroom tasks to transform reading input into audience-oriented output, with same-text collaboration reducing cognitive load before the return to different-text integration.

For the same reason, the development observed in the presentation scripts should not be attributed primarily to AI support; at least three alternative or complementary explanations must be considered. First, by the third stage, learners had repeatedly performed the same type of presentation task, and improved organization may partly reflect growing task familiarity. Second, the scaffolds that supported script development, such as vocabulary-reading practice and text-organization learning, were teacher-designed components, so improvement may reflect the staged task design as much as the AI-generated materials it delivered. Third, scripts were prepared through pair and group collaboration, so organizational improvements may represent negotiated group products and peer influence rather than purely individual development. These explanations cannot be disentangled within this design; the AI-generated materials are therefore best understood as one component of a broader instructional ecology in which AI generation, teacher scaffolding, task repetition, and peer collaboration operated together.

Although this staged design is specific to the present system, learners, and tasks, it offers an insight that other practitioners may examine in their own contexts. Reading-to-speaking development in this study required more than comprehension checks. Learners needed support in identifying core information, understanding text structure, judging information importance, and reorganizing information for listeners. Generative AI supported these processes here by providing vocabulary, grammar, text-organization, and comprehension scaffolds, but classroom interaction remained necessary for negotiation, audience awareness, and collaborative meaning-making.

Implications

The study offers three design insights. Derived from a single small-scale implementation, they are presented as transferable considerations for similar intermediate JFL settings, not as general prescriptions for JFL instruction. First, designing explicit mechanisms for prompt control and output evaluation proved more productive than ad hoc chatbot use, because it kept AI-generated content under instructional management. Second, learning logs were most informative when treated not merely as records of completion but as evidence of learner strategy, difficulty, and readiness for later tasks. Third, AI-supported individual learning became meaningful when it was linked to classroom activities that required learners to use, reorganize, and communicate what they had learned. For intermediate JFL instruction, a potentially useful design principle is to connect reading support with oral information-sharing tasks through vocabulary-reading practice and text-organization scaffolding.

Limitations and Future Research

This study has several limitations. First, it involved only five learners in one university context. In addition, the participants were self-selected volunteers who responded to an open recruitment notice, and they may therefore have been more motivated or more interested in AI-supported learning than their non-participating peers. The small, self-selected sample allowed detailed analysis of learner pathways, but the findings cannot be generalized to all JFL learners or all AI-supported environments. Second, the study did not include a control group. Observed changes may have resulted from the combination of AI-supported learning, classroom task revision, teacher support, repeated practice, and increased familiarity with presentation tasks. The study should therefore be understood as exploratory design research rather than causal evidence of effectiveness.

Third, the implementation period was short. Seven sessions were sufficient to observe changes in learning pathways and presentation-script organization, but longer-term research is needed to examine sustained development and transfer to other speaking tasks. Fourth, the teacher-researcher served as system designer, instructor, observer, and analyst. This made formative adjustment possible but may have influenced interpretation. All coding and pathway classification were conducted by the author

alone, without inter-rater reliability checks. Although coding followed Appendix C and interpretations were triangulated and checked with learners, the analysis still reflects a single researcher's perspective. Future studies should involve additional raters, coding discussion, and independent observers for classroom interaction. Fifth, although this study describes the teacher's quality-control procedure for AI-generated materials and reports readability scores, future research should evaluate the linguistic accuracy, difficulty appropriateness, and pedagogical alignment of AI output more systematically and with multiple evaluators. Sixth, learner-voice data remain thinner than system and log data. The mini-interviews were short formative check-ins; although episodes of learner doubt about AI answers and comments on ease and difficulty were captured, trust in AI-generated explanations and the emotional or cognitive dimensions of sustained system use were not systematically explored, and the account remains more system-centered than phenomenological.

Future research should test similar designs with larger and more diverse learner groups, compare different forms of AI-supported scaffolding, and investigate how teachers can use log data for formative assessment without overburdening their workload. It would also be valuable to examine learners' perceptions of AI feedback and speaking practice, especially the psychological barriers that may remain even when AI conversation tools are technically available.

Conclusion

This study examined how a generative AI-supported Japanese learning system could be integrated into intermediate JFL instruction to support personalized learning and oral information sharing. The system combined prompt-controlled content generation, immediate feedback, and learning-log storage within a backward-designed sequence that connected individual AI-supported reading with collaborative classroom presentation tasks.

The findings indicate that personalization occurred through teacher-controlled content adaptation, learner-regulated pace and sequence, and immediate feedback. Learning logs, interpreted together with classroom observations and mini-interviews, suggested iterative, stable-default, partial-compensatory, and expansive pathways of self-regulated learning. Presentation-script analysis showed that learners initially juxtaposed information from different texts, then stabilized information selection through same-text collaboration, and later began to reorganize information under emerging higher-order categories, showing partial but clear progress toward clearer oral information sharing.

The central contribution of the study is not to claim that generative AI alone improves language learning or that the observed changes generalize across JFL contexts. Rather, it offers a pedagogical design model showing how generative AI can become meaningful when embedded in a teacher-designed learning ecology. In this model, AI supports flexible materials generation, individualized preparation, feedback, and process visibility, while classroom activity supports collaboration, interaction, and communicative use. For similar intermediate JFL contexts, the study suggests one practical direction for using generative AI as an adaptable component of human-guided language learning.

Conflict of Interest

The author declares no potential conflicts of interest with respect to the research, authorship, and publication of this study.

Funding

This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Use of Generative AI

Generative AI was used in three distinct ways. First, the study itself examined a generative AI-supported Japanese learning system as the object of research; this system used GPT-4o for text generation and feedback, Whisper-1 for speech recognition, and TTS-1 for speech synthesis. Second, during manuscript preparation, ChatGPT (GPT-5.4 Thinking) was used to assist with translation from Japanese to English, language polishing, and structural revision. Third, Figure 4 was created with Claude Opus 4.8 based on the contents of saved JSON learning-log files in order to present the system's feedback output in a more readable form. The figure was checked against the original log data and revised by the author before inclusion. Generative AI was not used to generate the research questions, conduct coding, determine the findings, or draw conclusions. The author checked and revised all AI-assisted wording, interpretations, citations, visual presentation, and claims and take full responsibility for the manuscript.

References

- Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, Article 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- Chapelle, C. A. (2001). *Computer applications in second language acquisition: Foundations for teaching, testing and research*. Cambridge University Press.
- Choi, K. Y., Wu, C., & Moorhouse, B. L. (2025). Exploring the use of Generative Artificial Intelligence (GenAI) in English language teaching: Voices from in-service teachers at an early-adopting Hong Kong secondary school. *Technology in Language Teaching & Learning*, 7(3), 102516. <https://doi.org/10.29140/tltl.v7n3.102516>
- Hellmich, E. A., Vinall, K., Brandt, Z. M., Chen, S., & Sparks, M. M. (2024). ChatGPT in language education: Centering learner voices. *Technology in Language Teaching & Learning*, 6(3), 1741. <https://doi.org/10.29140/tltl.v6n3.1741>
- Ishiguro, K. (2020). *Danraku-ron: Nihongo no wakariyasusa no kimete [Paragraph theory: The key to clarity in Japanese]*. Kōbunsha.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Chen, D., Dai, W., Chan, H. S., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 248. <https://doi.org/10.1145/3571730>
- Kobayashi, N., Ford-Niwa, J., & Yamamoto, H. (1996). Nihongo nōryoku no atarashii sokuteihō: SPOT [SPOT: A new method of measuring Japanese proficiency]. *Japanese Language Education Around the Globe*, 6, 201–218. <https://doi.org/10.20649/00000239>
- Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.
- Lee, J. (2016). Nihongo kyōiku no tame no bunshō nan'ido kenkyū [Readability Research for Japanese Language Education]. *Waseda Nihongo Kyōikugaku*, 21, 1–16. <https://hdl.handle.net/2065/00051749>
- Li, B., Tan, Y. L., Wang, C., & Lowell, V. (2025). Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. *Computers and Education: Artificial Intelligence*, 9, 100445. <https://doi.org/10.1016/j.caeai.2025.100445>
- Liu, G., & Ma, C. (2024). Measuring EFL learners' use of ChatGPT in informal digital learning of English based on the technology acceptance model. *Innovation in Language Learning and Teaching*, 18(2), 125–138. <https://doi.org/10.1080/17501229.2023.2240316>
- OpenAI. (n.d.). *Best practices for prompt engineering with the OpenAI API*. OpenAI Help Center. Retrieved April 18, 2026, from <https://help.openai.com/en/articles/6654000-best-practices-for-prompt-engineering-with-the-openai-api>

- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). Academic Press. <https://doi.org/10.1016/B978-012109890-2/50043-3>
- Sakuma, M. (1999). Gendai Nihongo no bunshō kōzō ruikai [Typology of text structures in contemporary Japanese]. *Memoirs of the Japan Women's University. Faculty of Literature*, 48, 1–28.
- Stockwell, G. (2024). ChatGPT in language teaching and learning: Exploring the road we're traveling. *Technology in Language Teaching & Learning*, 6(1), 2273. <https://doi.org/10.29140/tltl.v6n1.2273>
- Stockwell, G., & Wang, Y. (2025). Framing human-AI collaboration in language education. *Technology in Language Teaching & Learning*, 7(4), 103824. <https://doi.org/10.29140/tltl.v7n4.103824>
- Strobl, C., Menke-Bazhutkina, I., Abel, N., & Michel, M. (2024). Adopting ChatGPT as a writing buddy in the advanced L2 writing class. *Technology in Language Teaching & Learning*, 6(1), Article 1168. <https://doi.org/10.29140/tltl.v6n1.1168>
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- University of Tsukuba. (n.d.). *Tsukuba Test-Battery of Japanese: TTBJ no kojūn jiken ni tsuite* [About individual testing]. Retrieved April 21, 2026, from <https://ttbj.cegloc.tsukuba.ac.jp/p1.html#SPOT>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wang, Y., Zhang, T., Yao, L., & Seedhouse, P. (2025). A scoping review of empirical studies on generative artificial intelligence in language education. *Innovation in Language Learning and Teaching*. Advance online publication. <https://doi.org/10.1080/17501229.2025.2509759>
- Warschauer, M. (1996). Computer-assisted language learning: An introduction. In S. Fotos (Ed.), *Multimedia language teaching* (pp. 3–20). Logos International.
- Wiggins, G., & McTighe, J. (2005). *Understanding by design* (Expanded 2nd ed.). ASCD.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2

Appendix A: Representative Prompt Excerpts

The following excerpts are representative prompts used to generate and revise learning components. The full prompt set can be provided as supplementary material if requested by the journal.

Table A1. *Representative Prompt Excerpts*

Learning component	Prompt excerpt translated from Japanese
Reading text generation	Create a Japanese reading passage about the specified topic. Target level: JLPT N3/CEFR B1. Length: 480-520 Japanese characters. Structure the passage as opening -> development -> closing. Use vocabulary and grammar appropriate to the target level. If vocabulary not included in the uploaded vocabulary list is used, provide a new-vocabulary list with readings and meanings. Output format: title, text, new-vocabulary list if applicable.
Vocabulary learning	Read the title and Japanese text. From words in the text not included in the vocabulary list, select 25 items that are valuable for intermediate learners. For each item, provide the word, reading, translate meaning, and usage explanation. Exclude words that are too elementary and prioritize vocabulary useful for text comprehension and expression. Balance native Japanese words, Sino-Japanese words, and loanwords where possible.
Vocabulary-reading practice	Select 15 important vocabulary items from the text for intermediate learners and create four-option questions asking for the reading of each item. Provide answers and explanations only after the learner responds. Include distractors that learners are likely to confuse.
Text-organization learning	Analyze the structure of the reading text so that learners can understand it more easily. Explain the role of each part and the overall development of the text in clear language.
Reading-comprehension practice	Create five four-option content-comprehension questions based on the reading text, following JLPT-style reading question formats. Questions should test main idea, key points, paragraph relationships, cause-effect, contrast, author intention, stance, and implication. Each question should have only one correct answer, and distractors should be plausible for learners who have not read the text carefully.

Appendix B: Codebook for Log Analysis

Table B1. *Log action Codes*

Repetition	the same practice component was attempted more than once within a session.
Return to foundational learning	a learner moved from a practice task back to vocabulary or grammar learning after receiving a low score, skipping an item set, or showing uncertainty.
Movement toward advanced practice	a learner completed the reading module and then entered the optional conversation-practice module.
Repair	a return to foundational learning, or a repetition, whose purpose of restoring understanding was confirmed by converging evidence (a preceding low score and a subsequent improved score, classroom observation of note-taking or note consultation, or the learner's own explanation in the mini-interview).

Appendix C: Codebook for Pathway Classification

Classification proceeded in two steps for each learner and session: (1) the logs were inspected for the presence or absence of repetition, returns to foundational learning, sequence changes, and entry into optional practice; (2) the reason for each identified pattern was confirmed directly with the learner in the mini interview before a label was assigned. No quantitative thresholds (e.g., minimum number of returns) were used; classification was qualitative and evidence-based.

Table C1. *Pathway Categories and Decision Rules*

Iterative/compensatory pathway	the learner repeatedly returned from practice tasks to foundational learning after low scores or self-reported uncertainty, with the repair purpose confirmed in the interview.
Stable-default pathway	the learner followed the system's default sequence without returns or repetitions, with stable scores indicating that the default sequence matched the learner's needs.
Partial-compensatory pathway	the learner mostly followed the default sequence but inserted local review or repetition at specific points of difficulty.
Expansive pathway	the learner completed the required components efficiently and moved into the optional conversation-practice module.

Where a learner showed features of more than one category, the dominant pathway was assigned on the basis of the most frequent and interview-confirmed pattern, and mixed cases are reported as such (e.g., repair plus task-oriented reordering for Participant C).

Appendix D: Codebook for Script Analysis

Every sentence in each AI-generated source text and in each learner presentation script was assigned a sequential number. Each numbered sentence was then coded for its semantic function within its paragraph and within the text as a whole (e.g., topic-presenting, explaining or elaborating, exemplifying, contrasting, transitional, summarizing or concluding). Information unit: a clause or sentence group conveying a distinct idea from the source texts. Retained information: a unit traceable to a numbered source-text sentence. Omitted information: a numbered source-text unit absent from the script. Reorganized information: a retained unit whose position or grouping in the script differed from the source text. Information selection and organization: whether important units were selected and arranged for listeners through grouping, sequencing, or categorization. Logical coherence: whether the script showed an opening, development, and conclusion, with relationships among units made explicit through topic sentences, connective expressions, and summary statements.