

Evaluating the Accuracy and Level-Appropriateness of ChatGPT Feedback for Beginning Korean Learners: Implications for AI-Assisted Language Teaching

Inho Jung

Brigham Young University-Provo

Abstract

The integration of artificial intelligence (AI) into language learning has introduced new opportunities for personalized instruction and real-time feedback. This study investigates the accuracy and level-appropriateness of ChatGPT's feedback for beginning Korean learners at Brigham Young University (BYU-Provo). Using the free version of ChatGPT 3.5 within the beginning level curriculum, the research examines how accurately and appropriately ChatGPT provides meaningful and effective feedback to students. Employing Feedback theory, the study collects data through interaction logs and surveys to assess the accuracy of ChatGPT's feedback, its impact on student confidence and engagement, and its role in fostering individualized learning experiences. The findings will provide insights into AI's feasibility as a supplementary language tutor, highlight the need for developing a customized GPT for specific learner groups and for training students how to create specific prompts tailored to their

learning goals. Additionally, this study explores how AI can assist teachers in providing effective language instruction.

Keywords: Teaching Korean; A.I.; ChatGPT; Feedback; Accuracy; Appropriateness

1. Introduction

The integration of artificial intelligence (AI) into education has rapidly shifted the ways in which learners access, engage with, and process language input. Among the most notable innovations is the use of generative AI tools like ChatGPT, which can produce natural language output and interact conversationally with users. These tools are gaining attraction in second language education for their capacity to simulate communicative contexts, provide personalized practice, and offer immediate responses tailored to learner input (Lee, 2024).

Such developments are particularly significant for beginning language learners, who often lack access to sufficient language exposure and individualized corrective feedback—both of which are essential for foundational development. In Korean language education, these challenges are amplified by the complexity of the language's honorific system, syntax, cultural vocabulary, and sociolinguistic norms, which require both precision and cultural sensitivity. While human instruction remains indispensable in this process, there is growing interest in exploring whether AI tools can function as supplementary teaching assistants to reinforce early-stage language acquisition.

Despite the increasing adoption of AI in educational settings, questions remain regarding the pedagogical quality of the feedback it provides. Specifically, there is limited empirical evidence addressing whether AI-generated feedback—especially from tools like ChatGPT—can meet the linguistic, contextual, and instructional needs of novice learners. This is a particularly relevant concern in Korean language education, where improper corrections or overly generalized explanations may impede rather than support learning progress.

This study seeks to fill that gap by evaluating the linguistic accuracy and pedagogical level-appropriateness of ChatGPT 3.5's feedback for beginning Korean learners. By analyzing and evaluating AI-generated feedback, this research aims to determine the extent to which ChatGPT can serve as a reliable and instructionally sound support tool in Korean language classrooms.

2. Literature Review

2.1. A.I. in Language Education

The integration of Artificial Intelligence (AI) in education has significantly influenced language learning by enabling personalized feedback, adaptive learning environments, and automated assessment tools. These technologies facilitate

real-time interaction and support learner autonomy, which are particularly beneficial in developing linguistic competence.

One of the most widely adopted AI features in language education is conversational AI. Tools like *EnglishBot*, introduced by Ruan et al. (2021), simulate human dialogue to provide learners with contextual and interactive speaking practice. This type of real-time, input-driven learning helps bridge the gap between structured classroom instruction and natural language use.

Nam (2023) reported that a multimodal AI chatbot named *Sayromi* significantly improved learner engagement and performance. Similarly, Belda-Medina and Calvo-Ferrer (2022) highlighted how AI chatbots serve as immediate feedback providers and conversational partners, supporting grammar and vocabulary development through contextual learning. Their findings suggest that learners benefit from the accessibility and responsiveness of AI tools, which complement traditional instruction. These benefits point to AI's potential to foster deeper cognitive involvement in language processing.

However, concerns have also been raised about the reliability and level-appropriateness of AI-generated feedback. Pikhart (2020) noted that while AI applications in language learning can efficiently detect errors, they may struggle with

handling culturally or pragmatically nuanced content. As AI tools like ChatGPT become increasingly present in classrooms, questions remain about their ability to offer high-quality feedback that aligns with human expert judgment.

In summary, AI technologies are reshaping language education by making learning more interactive, responsive, and tailored to individual needs. Nevertheless, a critical evaluation of how these tools deliver feedback—particularly for novice learners—is essential to ensure their pedagogical soundness and instructional reliability and appropriateness.

2.2. The Feedback Theory and Feedback of A.I. in Language Education

Feedback has long been regarded as a cornerstone of effective language learning. Hattie and Timperley (2007) defined effective feedback as that which addresses three fundamental questions: “Where am I going?” (learning goals), “How am I going?” (performance), and “Where to next?” (strategies for improvement). Their model emphasizes feedback as a tool for learning progression rather than simply correction.

Building on this foundation, Shute (2008) emphasized that formative feedback should be timely, goal-referenced, and cognitively accessible. She argued that effective feedback

not only ensures accuracy but also reduces cognitive overload—an area where AI-generated feedback may sometimes fall short, as it can provide information that exceeds the learner's current level of understanding.

A few studies have explored the instructional impact of AI-based feedback. Wang (2024) demonstrated that EFL students who received writing feedback from ChatGPT experienced significant gains in grammatical accuracy and reduced anxiety compared to those relying solely on teacher feedback. In broader applications, AI-supported platforms have also been shown to address emotional and motivational aspects of learning. For example, Yuan (2025) reported that an AI reading platform augmented with biometric feedback significantly improved learners' reading comprehension and motivation, while also reducing anxiety. This reinforces the value of AI in both cognitive and affective dimensions of language education. These findings suggest that AI can play a supportive role in helping learners internalize structural language knowledge.

Nonetheless, the literature suggests that while AI feedback can enhance linguistic accuracy, motivation, and learner engagement, concerns regarding the accuracy and level-appropriateness of AI feedback persist. For beginning Korean learners, these concerns are especially critical, as

early-stage learners are more susceptible to adopting incorrect or overwhelming input. Thus, evaluating the accuracy and level-appropriateness of AI-generated feedback is essential to leveraging its full potential in language education.

Research Questions:

1. To what extent are the corrections and feedback provided by ChatGPT accurate and supportive of beginning-level Korean learners' language development?
2. How appropriate and cognitively accessible is ChatGPT's feedback for beginner learners, based on their proficiency level?
3. What are the implications of the findings for developing a customized GPT and training students in effective AI usage?

By analyzing ChatGPT's performance in identifying and correcting linguistic errors and providing feedback specific to beginner Korean learners, this research will contribute valuable insights into its applicability as a language learning tool. The findings will inform educators and learners about the strengths and limitations of AI-generated feedback, providing essential data for optimizing AI-based instruction and determining whether a specialized GPT model is required to better serve Korean language learners or is enough after training them.

3. Methodology

3.1. Participants

The participants were 35 undergraduate students enrolled in two sections of Korea 101 at Brigham Young University–Provo. All students were novice-level Korean learners with limited prior exposure to the language. Participation in the study was voluntary, and all responses were collected anonymously.

3.2. Instruments

Data were collected through a custom-designed online survey administered via Qualtrics. The survey included Likert-scale items, multiple-choice questions, and open-ended prompts. It was designed to assess the following five dimensions:

1. Time spent and topics studied using ChatGPT for Korean learning.
2. Learning goals and types of prompts submitted to ChatGPT.
3. Students' perceptions of the accuracy of ChatGPT's feedback.
4. Reflections on the helpfulness, confusion, or suggestion elements in the feedback.
5. Teachers' evaluations of ChatGPT's feedback in terms of accuracy and level-appropriateness.

The 5-point Likert scale used to rate the accuracy and level-appropriateness of ChatGPT's feedback was developed with reference to Hattie and Timperley's (2007) feedback theory, which outlines that effective feedback should include feed up (learning goal), feed back (current performance), and feed forward (future improvement). Although these dimensions are conceptual, the study adapted them into a quantifiable scoring rubric to empirically measure feedback quality. Table 1 below draws primarily from Hattie and Timperley's (2007) feedback model, which evaluates feedback based on its ability to guide learners through the dimensions of learning goals (feed up), current progress (feed back), and next steps (feed forward).

Table 1. Accuracy Rating Scale

Score	Description	Feedback Theory Alignment
5	Very Accurate	Fully correct; includes feed up, feedback, and feed forward (Hattie & Timperley, 2007)
4	Somewhat Accurate	Minor flaw; includes at least one and two errors in feedback dimensions
3	Neutral	Mixed accuracy; task-level only; lacks depth or instructional direction, more than three errors in feedback dimensions
2	Somewhat Inaccurate	Mostly incorrect or vague; more than four errors in feedback dimensions. lacking actionable guidance or feedback structure
1	Very Inaccurate	Entirely incorrect; pedagogically misleading or unstructured

In addition to Hattie and Timperley’s (2007) feedback levels, the scale integrates Shute’s (2008) emphasis on feedback being timely, goal-referenced, and cognitively manageable to prevent learner overload. Each score level

reflects theoretical constructs of effective instructional support and practical appropriateness. Table 2 below is grounded in Shute’s (2008) principles of cognitive alignment to reduce overload and enhance comprehension for target learners.

Table 2. Level-Appropriateness Rating Scale.

Score	Description	Feedback Theory Alignment
5	Very Appropriate	Clear, level-matched, timely, and reduces cognitive overload (Shute, 2008)
4	Somewhat Appropriate	Generally clear with minor issues; mostly appropriate for learner level
3	Neutral	Occasionally misaligned or over-explained; moderate cognitive load
2	Somewhat Inappropriate	Lacks clarity; uses advanced terms or confusing phrasing
1	Very Inappropriate	Confusing or unstructured; poorly aligned with learner proficiency

This adaptation enables an evaluation of the extent to which ChatGPT’s feedback aligns with both linguistic

accuracy and pedagogical appropriateness, as informed by feedback theory. Student perceptions of feedback accuracy were also assessed using this 5-point scale to compare with those of the instructor and researcher. In addition, the instructor and researcher used the same scale to independently evaluate both accuracy and appropriateness.

3.3 Procedure

Students were encouraged to use ChatGPT 3.5, a free version, regularly to support their Korean learning throughout the semester. They interacted with the tool by asking questions related to vocabulary, grammar, speaking, reading, and writing. In Week 10, students completed the survey, where they reported their study time, the content of their interactions with ChatGPT, and their impressions of the feedback received. They also submitted examples of their prompts and ChatGPT's responses. Students were allowed to submit the survey more than once during the study period. Informed consent was obtained at the beginning of the survey, and all responses were kept anonymous for research purposes.

3.4 Data Analysis

Quantitative responses were analyzed using descriptive statistics, including frequency counts and percentages, to

determine trends in students' evaluations of feedback accuracy and level-appropriateness. Open-ended responses were coded thematically to identify patterns in perceived benefits, challenges, and suggestions for improving ChatGPT's feedback. These student responses were triangulated with instructor and researcher evaluations of the same ChatGPT outputs to compare learner perceptions with pedagogical assessments and strengthen the study's internal validity.

4. Results

4.1. ChatGPT Usage and Study Topics

A total of 52 responses were collected from 35 students who participated in the survey. Most students reported using ChatGPT for Korean language study for approximately 30 to 60 minutes per week. When asked to specify their study topics, many students selected more than one category, reflecting varied learning goals across linguistic domains. The most commonly selected areas of focus were grammar ($n = 19$), vocabulary ($n = 15$), and speaking practice ($n = 8$), followed by writing ($n = 6$) and reading ($n = 2$). These results suggest that students primarily used ChatGPT for form-focused activities such as grammar and vocabulary, while also exploring productive skills like speaking and writing to a lesser extent. Students engaged with ChatGPT in

diverse ways, often seeking task-level support such as particle usage, conjugation rules, and spelling, as well as more process-level feedback like interpretation and clarification of Korean texts. Example prompts included:

- “Can you help me with my spelling practice?”
- “What’s the difference between 예 and 예 사?”
- “Explain the -거 예 요 ending.”

These prompts illustrate students’ self-directed attempts to elicit both corrective and explanatory feedback—key components of effective learning support as described by Hattie and Timperley (2007).

4.2. Students’ Perceived Accuracy of Feedback

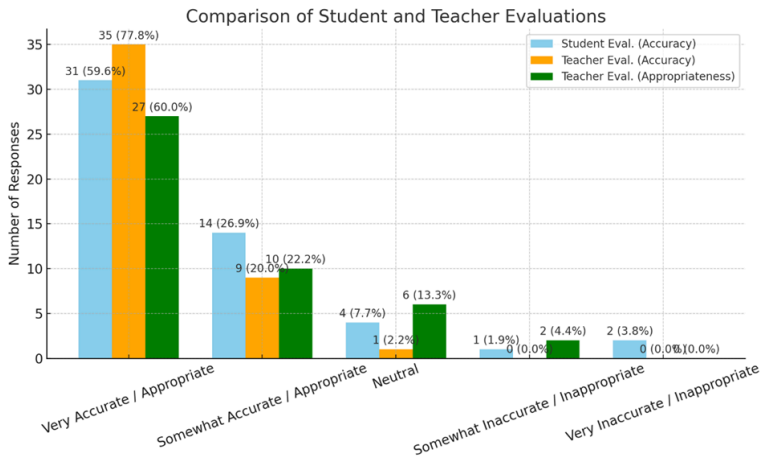
Following the rating criteria outlined in the Instrument section, student perceptions reflected a generally favorable view of ChatGPT’s feedback quality. Of the 52 responses, 60% (n = 31) rated the feedback as Very Accurate, suggesting it fully aligned with their expectations for task and process accuracy. 26.9% (n = 14) marked Somewhat Accurate, implying mostly correct content with a minor flaw. 7.7% (n = 4) selected Neutral, indicating multiple partial inaccuracies that disrupted clarity or confidence. 1.9% (n = 1) rated it as Somewhat Inaccurate, suggesting feedback lacked adequate

support to guide learning and 3.8% ($n = 2$) identified the feedback as Very Inaccurate. Although only 60% considered the feedback entirely accurate, the fact that 86.9% fell into the top two categories suggests a strong perceived instructional value. According to feedback theory, this likely reflects ChatGPT's effectiveness in providing "feedback" on performance and partial "feed forward" on next steps. The minority who rated it lower may reflect moments of misalignment with learning goals, or a lack of clarity in the feedback's instructional message.

4.3. Teacher and Researcher Assessment of Feedback

To validate the learners' self-assessments, the course instructor and an independent researcher evaluated 52 feedback examples submitted by students. These were assessed using a rubric grounded in feedback theory to evaluate both feedback accuracy and pedagogical level-appropriateness.

4.3.1. Teacher and Researcher Evaluation of ChatGPT’s Feedback



Teacher and researcher evaluations were conducted using the five-point scale to assess feedback accuracy and appropriateness. Results show strong endorsement of ChatGPT’s instructional quality:

- 77.8% (n = 35) of feedback was rated as Very Accurate (5 points)—meeting all core elements of effective feedback as outlined in Hattie and Timperley’s model.
- 20.0% (n = 9) were marked as Somewhat Accurate (4 points), reflecting only one minor inaccuracy.
- 2.2% (n = 1) received a Neutral score (3 points).
- None were considered inaccurate.

While the majority of ChatGPT's feedback was positively evaluated by both the teacher and researcher, the fact that only 77.8% of responses were rated as Very Accurate (5 points) warrants careful attention. Although this figure reflects strong overall performance, it also reveals that nearly one in four responses contained at least minor inaccuracies or lacked full alignment with established feedback criteria. In high-stakes language learning environments—particularly for beginners who may internalize incorrect input—this level of imperfection is potentially concerning. Given that none of the feedback was rated as entirely inaccurate, the results affirm ChatGPT's general instructional reliability; however, the presence of even minor flaws emphasizes the necessity for ongoing instructor mediation, critical student engagement, and potentially further refinement or customization of AI feedback tools. Teachers and curriculum designers must remain vigilant in reviewing AI-generated content and training learners to evaluate and question the responses they receive; this is necessary to avoid the risk of fossilizing subtle errors. On the dimension of appropriateness, which focuses on proper cognitive load and level-appropriateness in line with Shute (2008) we have the following results:

- 60.0% ($n = 27$) of feedback responses were rated as Very Appropriate (5 points).

- 22.2% (n = 10) were rated Somewhat Appropriate (4 points).
- 13.3% (n = 6) were marked Neutral (3 points).
- 4.4% (n = 2) as Somewhat Inappropriate (2 points).

While 60.0% of the responses were rated as Very Appropriate (5 points), this also means that 40% fell below the highest standard. Specifically, 22.2% were considered only somewhat appropriate, and 17.7% ranged from neutral to somewhat inappropriate. These results suggest that even when feedback is linguistically accurate, it may not always be accessible or well-matched to the cognitive needs of beginner learners. Given Shute's (2008) emphasis on reducing cognitive overload and aligning feedback with learners' developmental stages, this finding underscores the need for greater attention to clarity and level-appropriateness in AI-generated responses. Without adequate scaffolding or simplification, even accurate content may fail to support learning effectively—particularly for novices who are more vulnerable to confusion or misinterpretation. This reinforces the importance of training learners to interpret AI feedback. It also reinforces the potential benefit of developing adaptive AI tools that tailor output based, not only on correctness, but also on learner proficiency.

4.3.2. Comparison with Student Perceptions

Compared to teacher ratings, students demonstrated a slightly more varied perception. Among 52 responses, 59.6% ($n = 31$) rated ChatGPT's feedback as Very Accurate, and 26.9% ($n = 14$) as Somewhat Accurate—86.5% positive overall. However, 7.7% ($n = 4$) selected Neutral, and 6.3% ($n = 3$) fell into the Inaccurate range.

The students' perceptions of ChatGPT's accuracy, when compared to the teacher and researcher evaluations, reveal important insights into how learners interpret and rely on AI-generated feedback. While the majority of students (86.5%) rated the feedback positively—closely aligning with expert assessments—there are notable differences in the distribution of lower ratings that carry pedagogical implications.

Specifically, teachers and researchers found no responses to be fully inaccurate, while 6.3% of students did. This suggests that some students may misjudge accurate feedback as inaccurate, potentially due to misunderstanding the explanation, differences in terminology from classroom instruction, or limited metalinguistic awareness. Conversely, students also gave slightly fewer top ratings than the expert raters, indicating that some learners may have found the feedback less helpful or less clear, even when it was technically correct.

These discrepancies highlight a key issue: learners evaluate feedback not just by correctness, but by perceived clarity, usefulness, and alignment with their expectations. For beginning learners especially, the ability to assess the quality of feedback may be limited, leading to under- or overestimation of its instructional value.

In summary, the variation between student and expert ratings suggests the need for explicit instruction on how to interpret and evaluate AI feedback, as well as continued teacher mediation to bridge gaps in perception and ensure learners benefit fully from AI-assisted tools.

Together, these findings emphasize that while ChatGPT is a valuable and accessible source of language input, it should not be viewed as a stand-alone instructional tool. Its effectiveness increases significantly when used in conjunction with teacher guidance, which can help clarify nuanced content, resolve mismatches in terminology, and guide learners in critically engaging with AI-generated responses. Particularly at the beginner level, teacher moderation is essential to ensure that AI feedback supports rather than undermines learner development.

4.4. Examples of ChatGPT Feedback: Clarity, Accuracy, and Appropriateness

4.4.1. Very Accurate and Very Appropriate Feedback (5/5)

Student Question: “ending 아요 or 어요”

ChatGPT’s Response:

The endings -아요 and -어요 are both used in Korean to make a sentence polite and are typically used in casual or semi-formal situations. They are used with verbs and adjectives to indicate politeness without being overly formal. The choice between -아요 and -어요 depends on the vowel of the verb stem:

- -아요: Used when the verb or adjective stem ends in ㅏ or ㅑ.
- -어요: Used when the verb or adjective stem ends in any other vowel.

For example:

- 가다 (to go) → 가요 (since the verb ends in ㅏ)
- 먹다 (to eat) → 먹어요 (since the verb ends in ㅓ)

Would you like examples with specific words or a more detailed explanation?

This response was rated as both highly accurate and appropriately leveled. The structured explanation, use of common examples, and the offer of further clarification demonstrate strong alignment with beginner-level needs and Shute's clarity principles.

4.4.2. Somewhat Accurate but Very Appropriate Feedback (4/5 Accuracy, 5/5 Appropriateness)

Student Question: "How to count 24 people in Korean?"

ChatGPT's Response:

Twenty-four students requires the use of native Korean numbers + counter, because the number of people typically uses native numbers, and the counter for people is 명.

학생 스물넷 명 (twenty-four students)

➡ 학생 (students) + 스물넷 (24) + 명 (people)

Korean number expression rules for people:

The counter for people is 명.

Use native Korean numbers for the digits:

1 = 하나 → 한 명

2 = 둘 → 두 명

3 = 셋 → 세 명

4 = 넷 → 네 명

24 = 스물넷 → 스물넷 명

Examples:

학생 한 명 (one student)

학생 두 명 (two students)

학생 세 명 (three students)

학생 열 명 (ten students)

학생 스물넷 명 (twenty-four students)

Although the overall explanation was clear and appropriate for the learner's level, the teacher and researcher rated the response as 4 out of 5 for accuracy. The specific phrase 학생 스물넷 명 is incorrect; in natural usage, it should be 학생 스물네 명, with the "ㅈ" dropped in compound native numerals. Fortunately, the student detected this discrepancy, noting, “different from our textbook, on our textbook it is ‘네 명.’”

This incident underscores a critical concern: when such subtle grammatical errors go unnoticed, they may lead to fossilization in the learner's interlanguage. It highlights the importance of teacher intervention and the potential risk of overreliance on AI-generated feedback.

4.4.3. Level-Inappropriate Feedback (3/5)

Student Question: “Can you help me with grammar in Korean? Verb conjugation and particles?”

ChatGPT’s Response (Summary)

The response offered a detailed and linguistically accurate explanation covering various grammatical categories, including present, past, and future tense conjugation rules, as well as the use of multiple particle types (e.g., subject, object, and location markers). Although the explanation was comprehensive, it was not tailored to the student’s proficiency level. The inclusion of advanced terminology and multiple grammatical concepts in a single response—such as tense distinctions and particle functions—may overwhelm beginning learners and hinder effective language acquisition. This suggests that general, open-ended questions may trigger ChatGPT to prioritize completeness and accuracy over level-appropriateness, highlighting the need for both student training in question formulation and AI calibration for pedagogical alignment. In fact, the student noted, *“I do think it gave more info than what I need for the course level right now.”* Another student similarly commented, *“I think I need to be a bit*

more specific in the question, so that ChatGPT doesn't overexplain or go into topics I wasn't really wanting to ask about.”

This example illustrates a recurring pattern: general or broad questions tend to elicit overly complex responses from ChatGPT. This suggests a need for both student training in how to formulate targeted prompts and AI calibration to ensure pedagogical alignment, especially for novice learners.

4.4.4. Implications from Students’ Questions and Feedback

1. Teach students how to ask effective questions that elicit accurate and level-appropriate feedback. Developing metacognitive awareness about question formation can improve the quality of AI-generated responses.
2. Train AI systems to adapt and calibrate feedback based on the specificity and clarity of learner questions. Enhanced sensitivity to input type can make AI responses more pedagogically aligned.
3. Educate students on the limitations of AI, emphasizing that AI-generated answers may not always be accurate. Encourage critical thinking and validation through teacher guidance.

4. Recognize that AI feedback remains valuable and useful, especially when used as a supplementary tool to reinforce learning, even with its imperfections.

5. Develop or adopt customized AI models tailored to specific learner proficiency levels, or advance adaptive AI systems capable of adjusting feedback dynamically to match the learner's current stage.

4.5. Qualitative Insights from Student Reflections

4.5.1. Helpfulness

Students appreciated ChatGPT's:

- Clear grammar breakdowns
- Personalized suggestions
- Interactive quizzes and practice prompts

One student commented, "It broke it down piece by piece. That made it so much easier to understand."

4.5.2. Limitations

Some noted:

- Over-explaining beyond their curriculum level
- Occasional misalignment with classroom terminology or instruction

One student said, “different from our textbook, on our textbook it is 네 명.”

4.5.3. Suggestions for Improvement

Some students recommended:

- More tailored, concise explanations
- Built-in follow-up questions
- Closer alignment with course level.

One student remarked, “i do think it gave more info than what i need for the course level right now, but it was helpful regardless..”

5. Discussion

This study explored the accuracy and appropriateness of ChatGPT 3.5's feedback for beginning Korean learners, drawing on triangulated data from student surveys and expert assessments. The findings affirm ChatGPT's instructional potential, particularly in providing immediate, task-level feedback that supports grammar, vocabulary, and sentence formation. Both teacher and researcher evaluations found that the majority of feedback was highly accurate (77.8%) and appropriately leveled (60%), suggesting that ChatGPT often meets the standards of effective formative feedback as described in Hattie and Timperley's (2007) and Shute's (2008) models.

However, the results also highlight key limitations that must be addressed when implementing ChatGPT in novice language learning contexts. Despite the high ratings for accuracy, approximately one in five responses contained minor inaccuracies or lacked full alignment with instructional expectations. This is a noteworthy concern in early-stage language acquisition, where subtle errors—particularly those involving particles, conjugation rules, or culturally embedded expressions—can be difficult for students to detect and may lead to the internalization of incorrect forms.

In terms of level-appropriateness, 40% of feedback fell below the “Very Appropriate” threshold, with some responses rated as neutral or somewhat inappropriate. These cases typically involved overly dense explanations, advanced terminology, or insufficient attention to learners’ cognitive load. Such mismatches suggest that even when content is correct, its delivery must be carefully calibrated to the learner’s developmental stage in order to be pedagogically effective.

Furthermore, a comparison of student and teacher/researcher assessments revealed perceptual gaps. While students generally rated the feedback positively, their evaluations were more variable and occasionally inconsistent with expert assessments. Some students misjudged correct

responses as inaccurate, while others failed to recognize subtle errors. This discrepancy highlights the influence of learner confidence, metalinguistic awareness, and familiarity with classroom instruction in shaping perceptions of feedback quality.

Taken together, these findings suggest that ChatGPT can serve as a valuable supplementary feedback tool, but its optimal use depends on several key conditions: learners must be trained to formulate clear and specific prompts; instructors must actively mediate and validate AI-generated responses; and AI systems themselves must be calibrated to provide level-appropriate, cognitively accessible feedback. Without these safeguards, there is a risk that ChatGPT may overwhelm or mislead learners, particularly those at the beginning stage.

In sum, while ChatGPT demonstrates strong potential as a support tool in language education, its effectiveness is maximized only when used in tandem with guided instruction and critical engagement strategies. These findings underscore the importance of integrating AI thoughtfully—balancing its capabilities with the pedagogical expertise of instructors and the learning needs of students.

6. Pedagogical Implications

6.1. Guided Prompt Bank for Effective ChatGPT Use

One key implication is the need for prompt engineering guidance—teaching students how to formulate specific, level-appropriate questions when using ChatGPT. Students in this study frequently received over-detailed responses, likely due to vague or open-ended prompts. To mitigate this, instructors can provide a “Prompt Bank” of model questions that demonstrate effective query formulation.

Examples of prompt templates include:

- Grammar clarification:
“Can you explain the difference between -에 and -에 서 with two beginner-level examples?”
- Conjugation help:
“How do I conjugate the verb 듣다 in the present tense using polite speech?”
- Vocabulary quiz practice:
“Give me a 5-question multiple-choice quiz on Korean school-related vocabulary.”
- Correction with reasoning:
“Here’s a sentence I wrote in Korean: ____ . Can you correct it and explain what was wrong at a beginner level?”

By incorporating prompt design practice into classroom activities or homework, students will not only receive more targeted feedback but also develop meta-cognitive skills in interacting with AI tools.

6.2. Student Checklist for Evaluating ChatGPT Responses

Another practical strategy is the integration of a student self-evaluation checklist or rubric to help learners critically assess the accuracy and appropriateness of ChatGPT's feedback. This encourages learner autonomy and prevents blind reliance on AI outputs, a concern raised in both student and teacher reflections.

Sample Student Checklist (Yes/No or 1–5 Scale):

- Is the explanation clear and aligned with what I have learned in class?
- Does the grammar/vocabulary usage match the lesson level (e.g., polite vs. informal speech)?
- Are the examples helpful and easy to understand?
- Do I recognize any information that conflicts with the textbook or instructor guidance?
- Does the feedback help me improve or understand what to do next?

This checklist could be distributed as a printable form, embedded in LMS platforms, or incorporated into learner reflection journals. It trains students to engage in critical AI literacy, a vital skill for 21st-century language learners.

6.3. Role of Instructor Mediation

Given that students may lack the linguistic awareness to discern nuanced inaccuracies in ChatGPT responses, teachers should act as mediators who validate, refine, or contextualize AI feedback. This can be achieved through:

- Weekly “ChatGPT Clinic” sessions where students submit AI feedback they received and discuss it in class.
- Annotated examples of both good and problematic AI feedback to analyze as a group.
- Instructor-curated prompts uploaded into a shared repository to model optimal AI use.

Such mediation ensures that AI-generated content becomes an instructional scaffold, not a source of confusion or misinformation.

6.4. Integration into Curriculum Design

To fully capitalize on the benefits of ChatGPT while minimizing potential drawbacks, its use must be embedded

within a broader instructional framework. This includes aligning AI engagement with course objectives, assessment strategies, and existing learning materials. For example, ChatGPT-assisted activities can be explicitly linked to units covering specific grammar points or vocabulary sets. Instructors may also assign weekly ChatGPT-based reflection logs, where students document their queries, summarize the AI's responses, and evaluate their usefulness based on classroom learning. Such integration not only reinforces target content but also promotes reflective learning practices. Additionally, course syllabi can designate ChatGPT as a “supplemental feedback partner,” with clear guidelines on its use, limitations, and the importance of verifying AI-generated information with teacher input or textbook references. This approach can empower students while maintaining instructional integrity.

6.5. Implications for Teacher Training and AI Literacy

These pedagogical strategies also underscore the evolving role of language instructors in the age of AI. Teachers must now be equipped not only with content expertise but also with a working knowledge of AI tool capabilities, prompt engineering principles, and digital literacy instruction. Professional development workshops on AI-mediated instruction, including how to guide students in crafting

effective prompts and critically analyzing ChatGPT's responses, would be instrumental. Moreover, developing institutional policies or department-level best practices around AI use in language classrooms can provide consistency and protect instructional quality. Instructors should also be encouraged to contribute to ongoing research and feedback loops that monitor the educational outcomes of AI integration.

6.6. Summary

Together, these implications reflect a shift in pedagogical design where AI tools like ChatGPT become embedded, scaffolded resources rather than standalone solutions. While ChatGPT shows strong potential as a real-time, personalized learning aid, its pedagogical value is maximized when students are trained to use it effectively and critically, and when instructors serve as mediators who contextualize its output. These practices align with feedback theory's emphasis on clarity, relevance, and learner engagement, reinforcing the idea that AI must be integrated thoughtfully to truly enhance second language acquisition.

7. Conclusion

This study contributes to the growing body of research on generative AI in language education by empirically examining both student perceptions and expert evaluations of ChatGPT 3.5. The findings indicate that students generally perceive ChatGPT as a valuable tool, particularly for receiving accurate and level-appropriate feedback in areas such as grammar, vocabulary, and sentence construction.

However, the analysis also reveals a critical need to calibrate AI-generated feedback to better align with beginner-level proficiency and classroom instruction. Subtle inaccuracies and overly complex explanations highlight the importance of adapting feedback to the learners' cognitive and instructional needs.

The triangulated results suggest that ChatGPT can function as an effective supplementary language learning assistant when used with appropriate scaffolding and instructional oversight. As generative AI becomes increasingly integrated into educational environments, its pedagogical value will depend on thoughtful implementation—balancing the strengths of AI with the essential guidance of human educators.

References

- Abbasi, K. J., Lashari, A. A., & Golo, M. A. (2025). Use of artificial intelligence as an online platform for English language teaching at the tertiary level. *Policy Journal of Humanities and Social Sciences*, 3(1), 1–12. <https://doi.org/10.5281/zenodo.14921342>
- Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, 12(17), 8427. <https://doi.org/10.3390/app12178427>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Lee, H. (2024). Perceptions, integration, and learning needs of ChatGPT among EFL teachers. *Korea TESOL Journal*, 20(1), 3–20.
- Ma, X., Zhang, C., & Lee, I. (2024). Learner engagement with AI-generated feedback among Chinese EFL students on second language writing. SSRN. <https://ssrn.com/abstract=5197066>
- Nam, S. (2023). The effect of using an intelligent, multimodal chatbot for Korean language education. *University of Hawaii ScholarSpace*. <https://scholarspace.manoa.hawaii.edu/items/8fd98b46-655d-41f0-9d68-776574a1e7d1>

- Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186X.2024.2410101>
- Pikhart, M. (2020). Intelligent information processing for language education: The use of artificial intelligence in language learning apps. *Procedia Computer Science*, 176, 2345–2352. DOI: [10.1016/j.procs.2020.09.151](https://doi.org/10.1016/j.procs.2020.09.151)
- Ruan, S., Jiang, L., Xu, Q., Davis, G. M., Liu, Z., Brunskill, E., & Landay, J. A. (2021). EnglishBot: An AI-powered conversational system for second language learning. In *Proceedings of the 26th International Conference on Intelligent User Interfaces (IUI '21)* (pp. 438–449). Association for Computing Machinery. <https://doi.org/10.1145/3397481.3450648>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Wandel, N., Stotko, D., Schier, A., & Klein, R. (2025). PyEvalAI: AI-assisted evaluation of Jupyter Notebooks for immediate personalized feedback. *arXiv preprint arXiv:2502.18425*. <https://arxiv.org/abs/2502.18425>
- Yuan, L. (2025). Artificial intelligence in language learning: biometric feedback and adaptive reading for improved comprehension and reduced anxiety. *Humanities and Social Sciences Communications*, 12(1), 1–12. <https://doi.org/10.1057/s41599-025-04878-w>

Zou, B., Huang, Y., & Xie, H. (2023). AI feedback in language learning: Complementing human instruction. *Computer Assisted Language Learning*, 36(4), 345–362. <https://doi.org/10.1080/09588221.2022.2067805>

Appendix

Customizing GPT for Practical Classroom Use

How to design a customized GPT

Simple Step-by-Step Guide to Create ‘BYU Korean Tutor’
GPT for Beginner Learners

Step 1: Set Your Goal

Create a GPT that helps beginner-level Korean learners (Korean 101/102) with vocabulary, grammar, expressions, and culture based on your own class materials.

Step 2: Go to ChatGPT and Start Creating

1. Open <https://chat.openai.com> and log in with a ChatGPT Plus account.
2. Click “Explore GPTs” and then click “Create a GPT.”
3. Choose “Walkthrough Mode” for a step-by-step guide.

Step 3: Name and Describe Your GPT

- Name: BYU Korean Tutor
- Short Description: An AI tutor designed to help beginning Korean learners at BYU-Provo.

- Long Instructions: Write what the GPT should do.
Use simple language like:

You are “BYU Korean Tutor,” a helpful assistant for beginner-level Korean learners at BYU. Refer to the textbook only. Use polite and simple language. Do not give information that is not in the textbook.

Step 4: Upload Textbook Materials (Important for Accuracy) in “Knowledge” to make sure the GPT gives accurate and level-appropriate answers:

- Upload the Korean 101 syllabus.
- Upload the textbook in Word or pdf document.
- Upload any handouts or worksheets you use in class.

This helps the GPT only use trusted material from your course.

Step 5: Add Example Questions and Answers You can type a few sample interactions to show the GPT how to respond:

- What does -고 있어요 mean?
- How do I say “I am eating” in Korean?
- Can you make a practice question with ‘입었어요?’

Step 6: Save and Share Your GPT

- Click “Publish”
- Copy the link and share it with your students through Canvas or QR code

Optional Tips

- Use it in class as a virtual assistant
- Ask students to practice with it before speaking tests
- Gather feedback and update it anytime (no coding needed)