https://dx.doi.org/10.20880/kler.2023.58.5.45

Learners' Experiences with Artificial Intelligence (AI) Chatbot for Korean Grammar Education

: A Focus on Usage Patterns and Perception

Song, Changkyung
Seoul National University
Department of Korean Language Education
Doctor course completion (1st author)

Seo, Heeju
Seoul National University
Department of Korean Language Education
Master's course student (2nd author)

Kim, Hojung
Seoul National University
Department of Korean Language Education Professor /
Korean Language Education Research Institute
Adjunct researcher (Corresponding author)

* This study was undertaken following the development of a multi-modal AI agent-based customized teaching and learning support system, with the backing of Seoul National University's College of Education in 2022 (Project number 700-20220011).

** An earlier version of this paper was presented at the 23rd International Conference on Education Research, Seoul, Korea.

I. Introduction
II. Literature Review
IV. Research Method
V. Results
VI. Discussion and Conclusion

I. Introduction

A chatbot is a machine agent that can engage in conversations with users through natural language, including both text and voice (Brandtzaeg & Følstad, 2018). Advances in natural language processing (NLP), machine learning, and deep learning techniques have led to the proliferation of artificial intelligence (AI) chatbots, which are widely used for various purposes, including customer service, technical support, coaching, communication, entertainment, and education in areas ranging from commerce to healthcare (Fidan & Gencel, 2022).

In an educational context, chatbots are utilized not only as conversational interlocutors but also as assistants to educators and partners for learners, stimulating learners' interest and guiding them in teaching and learning (Chen et al., 2020). They can function as interactive and engaging learning media by offering various materials to enhance the learning experience (Kharis et al., 2022).

Given this background, educational chatbots are becoming increasingly mainstream. Researchers have employed chatbot builders to develop context-specific chatbots in classrooms and examine their teaching effects (Ayedoun et al., 2015; Goda et al., 2014; Kim, 2016;

Kim, 2018a; Kim, 2018b; Wang et al., 2017). However, a meta-analysis of language education chatbots conducted by Huang et al. (2022) indicated that most studies focused on English as a Foreign Language (EFL) education and demonstrated a bias toward specific languages. In South Korea, chatbots for language education are designed for English education (Kim, 2016; Kim, 2018a; Kim, 2018b), and Korean as a foreign language (KFL) education (Park & Lee, 2021; Paik, 2021; Suh, 2021; Yoo & Yoo, 2021).

Research on chatbots targeting Korean language education among native speakers is limited, particularly among middle school students, as evidenced by a case study of elementary students (Min et al., 2020). However, given the wide range of educational stages from early childhood to lifelong learning, the perception of the educational benefits and usability of chatbot usage may vary with age. Accordingly, this study was conducted across different age groups.

This study highlights the importance of chatbots in normative grammar education for native youth speakers. The reluctance to take initiative in class stems from the extensive scale of normative grammar, which poses a challenge to its incorporation into the curriculum. Students may hesitate to ask questions, presuming that their middle school peers as native speakers possess an inherent grasp of their native grammar. They may also feel apologetic for their perceived lack of tacit knowledge (Polanyi, 1962).

Hence, this study explores the potential for middle school students to learn normative grammar through a chatbot that can consistently receive learners' queries both within and outside the school environment. Such a chatbot can reduce learners' anxiety and embarrassment toward initiating questions for teachers or peers (Davis, 2022, p. 38). The specific research questions are as follows:

- How do middle school learners employ and perceive chatbots for Korean grammar education?
- What are the implications for chatbot development, education using

chatbots, and chatbot research that may be adduced from this?

II. Literature Review

1. Chatbots for language learning

The use of chatbots for educational purposes can be traced back to educational agents in digital learning environments, known as intelligent tutoring systems, in the 1970s (Smuntny & Schreiberova, 2020). Traditional rule-based chatbots are based on a set of predefined guidelines extracted from external knowledge; thus, they are not very "intelligent" and cannot answer questions that they were not programmed to answer (Kohnke et al., 2023). However, owing to the development of advanced artificial intelligence (AI) technologies such as natural language processing (NLP), machine learning (ML), and deep learning (DL), chatbot technology has developed significantly. For example, Siri and Alexa, using ML, provide information and execute specific tasks upon request (e.g., turn on the radio, check the weather forecast), whereas OpenAI's ChatGPT is designed to engage in back-and-forth conversations with users (Kohnke et al., 2023). In addition to ChatGPT, large language models (LLM) such as Microsoft's BingChat, Google's Bard, and Anthropic's Claude, which most recently expanded their context window to 100,000 tokens (Park, 2023), compete in the AI-powered chatbot market.

Due to these rapid developments, AI-based chatbots are currently employed in various educational domains, including programming education, Chinese language learning, English language learning, mathematics education, and early childhood education (Fidan & Gencel, 2022). The educational purpose of chatbots in the context of language learning is first, to provide consistent support to students; second, to provide language information on various grammatical ex-

pressions and vocabulary that peers with similar proficiency cannot give; and third, to liberate humans (particularly teachers) from repetitive tasks, such as question-and-answer and practice (Huang et al., 2022, p. 238). Personality and emotional expressions, along with conversation records from log data, have contributed to the growing popularity of chatbots in language education (Park & Lee, 2021).

Chatbots can be classified according to various criteria, such as structure and purpose (Haristiani, 2019; Lee et al., 2019), modality (Park & Lee, 2021), operating method (Yoo & Yoo, 2021), and number of dialog contexts (Kim et al., 2019). Chatbots are classified by structure as tree-based flow chatbots with fixed responses, artificial intelligence chatbots that can continuously update their knowledge, or hybrid chatbots in an intermediate position. By purpose, they can be further classified based on specific objectives and functionalities, those designed for amusement, and those intended to mimic human conversations.

Chatbots can also be classified based on modality as text- and voice-based chatbots (Park & Lee, 2021), and by their operating methods into rule-based and machine-learning-based chatbots (Yoo & Yoo, 2021). Based on the number of dialog contexts, chatbot models can be classified as single- and multi-turn (Kim et al., 2019). Although the multi-turn chatbot discussed by Kim et al. (2019) differs from the chatbot developed in this study in that it refers to the full dialog history, it is common for single-turn to involve a short and safe response, whereas multi-turn is more natural and informative. This will be explained later; however, it is related to the reasons for the development of a chatbot driven by the system.

Meanwhile, the conversational abilities of most chatbots remain rule-based, relying on predefined scenarios or rule matching for specific databases, as they have not yet reached the level of natural, human-like communication (Lee et al., 2019, p. 133). Consequently, chatbots exhibit weaknesses related to low accuracy and feedback due to technical limitations, shortcomings in human connection and

accountability, privacy, and security issues (Davis, 2022), and challenges related to cognitive load and the novelty effect (Huang et al., 2022).

Therefore, the development of chatbots and their user experience enhancement relies on addressing technical issues and interaction design (Følstad & Brandtzaeg, 2020). Therefore, a post-development evaluation is necessary. Huang et al. (2022) provided criteria for analyzing the usefulness of chatbots in language learning contexts by dividing usefulness into usability and utility. Usability assesses whether the system efficiently and effectively accomplishes a series of tasks to satisfy users and is linked to technical affordances, whereas utility pertains to the educational and social affordances of the tool. Leveraging these three affordances can facilitate our investigation of the usefulness of educational chatbots in language learning contexts. A schematic of the criteria presented by Huang et al. (2022) is shown in Figure 1.

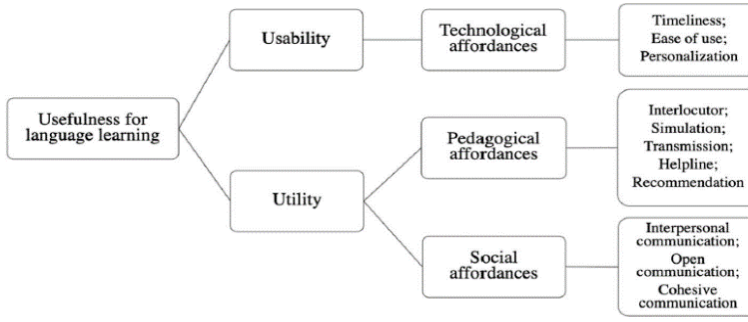

Figure 1. Affordances of chatbots in language learning (Huang et al., 2022, p. 253)

2. Korean normative grammar education

According to Min (2003, pp. 23-25), Korean language norms can be divided into narrow and broad categories. In a narrow sense, Korean language norms refer to four rules (Orthography, Pronunciation,

Loanword orthography, and Romanization). Broadly, they encompass the concepts of proper pronunciation, correct word usage, accurate sentence construction, appropriate conversational skills, and proper style in the written text. In this study, our discussion is based on the aforementioned broad sense; however, we limit the definition of Korean language norms to orthography and pronunciation.

Educational content related to normative language use in real life is dealt with in grammar subjects beginning in the 6th Korean language curriculum and to date constitutes a major component of grammar education. In particular, as achievement standards related to language norms explicitly began to be incorporated into the common curriculum during the 2007 revised curriculum period, Korean language norm education gradually began to be systematically established within grammar education (Nam, 2014). Table 1 presents the achievement standards related to Korean language norms during the curriculum revision period.

Table 1. Achievement standards related to Korean language norms

[2007, G10]	Know and use the Romanization and Loanword orthographies accurately
[2009, G7-9]	Understand the basic principles and contents of normative grammar
[2009, Korean I]	Learn the principles and contents of orthography and learn about cultured life with orthography
[2009, Korean II]	Understand the Romanization and Loanword orthographies and use them in real life
[2015, G7-9]	Pronounce and write the words correctly
[2015, G10]	Understand the basic principles and contents of orthography
[2022, G7-9]	Understand the basic principles and contents of orthography and apply them in real life

* 2007, 2009, 2015, 2022: Curriculum revision timing
* G7-9, G10: 1st to 3rd years in middle school, 1st year in high school.
* Korean I and II: subject names

One goal of Korean grammar education is to enable accurate language use, thereby inculcating learners' ability to engage in precise language activities. This is an important objective of grammar educa-

tion, which aims to provide knowledge and practice systematically through normative grammar education (Nam, 2014). Therefore, in this study, we narrowed down the scope of the chatbot development area to the Korean orthography and pronunciation covered in the [2015, G7-9] curriculum, as it targets third-year students of middle school who are completing the 2015 revised Korean language curriculum.

To apply the principles of normative grammar education to chatbot development, it is essential to first define the respective grammatical abilities. Purpura (2004, p. 86) viewed this as a combination of grammatical and strategic knowledge. Lee (2008) classified grammatical ability into knowledge, usage, and attitude. Koo (2010) defined it as the ability to understand and use Korean accurately and efficiently and further subdivided it into knowledge, the ability to explore, and the ability to apply knowledge to language activities. Yi (2013) defined grammatical ability as the ability to understand grammar, make grammatical judgments, and apply this understanding to various language experiences. Kim et al. (2007) and Nam (2007) introduced comprehension and application as subcomponents of grammatical ability.

Although the terminology varies among researchers, grammatical abilities go beyond mere propositional knowledge and encompass the ability to use and apply knowledge. In other words, in current grammar education, the emphasis is not only on comprehension but also on addressing the ability to explore phenomena, language use, and attitudes (Yu, 2017). The objective of the chatbot used in this study was to enable learners to acquire the knowledge of normative language independently through the inquiry into and application of language norms. Therefore, the questions provided by the chatbot to learners are organized by an "understanding" of the language regulation itself and "applying" an understanding of language regulations to real-life cases.

III. Research Method

1. Participants

A total of 62 participants were recruited from a middle school in Seoul, South Korea using convenience or opportunity sampling (Dörnyei, 2003, p. 72), whereby participants are selected based on specific practical criteria, such as geographical proximity, availability during certain time slots, and ease of access. In this case, participants were selected "purposively" to align with the research objectives. As the chatbot developed for this study was in the pre-commercialization stage, the available target experimental group was limited to the researcher's classroom. The specific characteristics of the participants are presented in Table 2.

Table 2. Characteristics of participants (N = 62)

Spec.		***N*(%)**
Gender	Male	34(54.8)
	Female	28(45.2)
Age	16 years old	62(100)
Spec.		***M*(*SD*)**
Korean grades (of 1 semester)		84.57(10.87)

2. Procedures

In this study, a chatbot for grammar education designed for middle school students was developed[1] based on the principles of

1 Part of Seoul National University, ETRI/TutorousLabs, Timbel Collaboration Project named Development of a Multimodal AI Agent-Based Learner-Customized Teaching and Learning Support System. In the chatbot development stage, data collection was

normative grammar education proposed in the research discussed in Chapter 2. The procedural timeline is as follows:

From March to June 2022, student queries and teacher responses were collected and supplemented with query-response data artificially generated by the researchers. The data were used to create a prototype Q&A chatbot that answered students' queries between June and August 2022. For H1 2023, additional grammatical query data were collected using the developed Q&A chatbot. In August 2023, a system-driven chatbot prototype was developed in which the chatbot initiated queries. Subsequently, the system-driven chatbot was pilot-tested with the participants for two weeks in September 2023.

The usage patterns and perceptions of the two chatbots were investigated through a survey conducted using a written questionnaire. A researcher (teacher) administered the questionnaires. Because the survey participants were young, parental consent was obtained before the survey, and the surveys were administered to groups of 30 students at a time (group administration). The survey duration was limited to 15-20 minutes to ensure that it did not exceed 30 minutes (following the principle outlined in Dörnyei, 2003, p. 18). A procedural flowchart is shown in Figure 2.

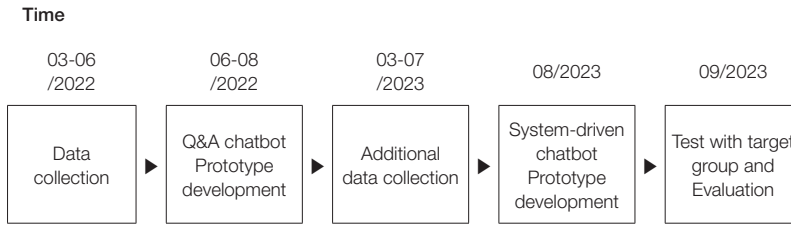

Figure 2. Research procedures

done by Seoul National University, engine development was done by ETRI/Tutorial Labs, and system development was done by Timbel. Therefore, in this paper, only the data collection stage was specified, and IRB was carried out by the project.

3. Instruments

1) Chatbots

Two types of chatbots were developed in this study: Q&A- and system-driven. Categorization depends on whether the student initiates the query or the chatbot takes the lead. Both are personified under the name “SNU Sammul.”

First, the Q&A chatbot is based on KoBERT,[2] which enables computers to understand and process text efficiently. If a student asks a question first, the chatbot responds based on the similarity between the newly entered queries and queries within the learned dataset. The user interface for the desktop version is shown in Figure 3.

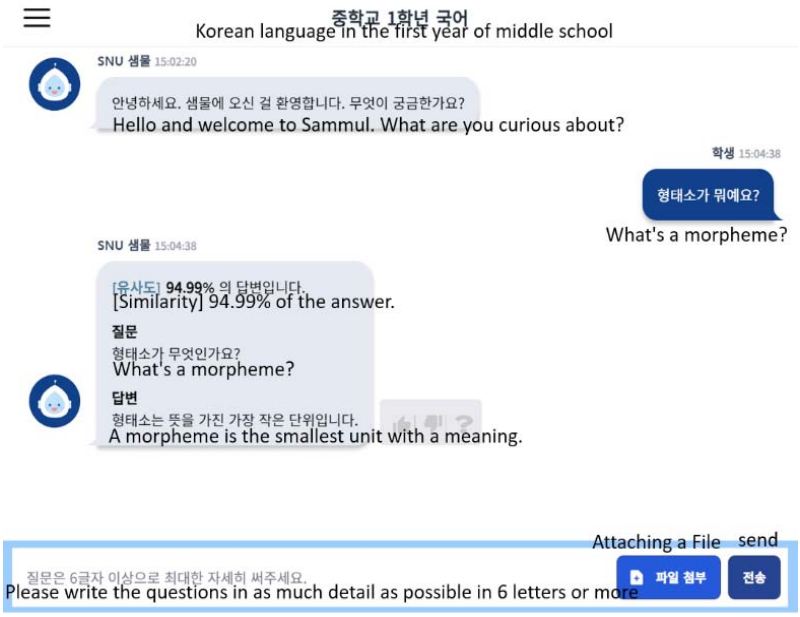

Figure 3. The user interface of the Q&A Chatbot

2 Word Embedding is a method of converting natural language so as to be understood

Unlike the Q&A chatbot, the system-driven chatbot is not a foundation model, but a newly built model in which the chatbot asks a question first, and the conversation proceeds through up to three turns depending on the student's response. To this end, the chatbot's dialog flow was set in advance following the principle that if the learner answered the question correctly, positive feedback and a final integrated explanation were provided; if the answer was wrong, the problem was explained, and other related questions were presented. This is illustrated in Fig. 4.

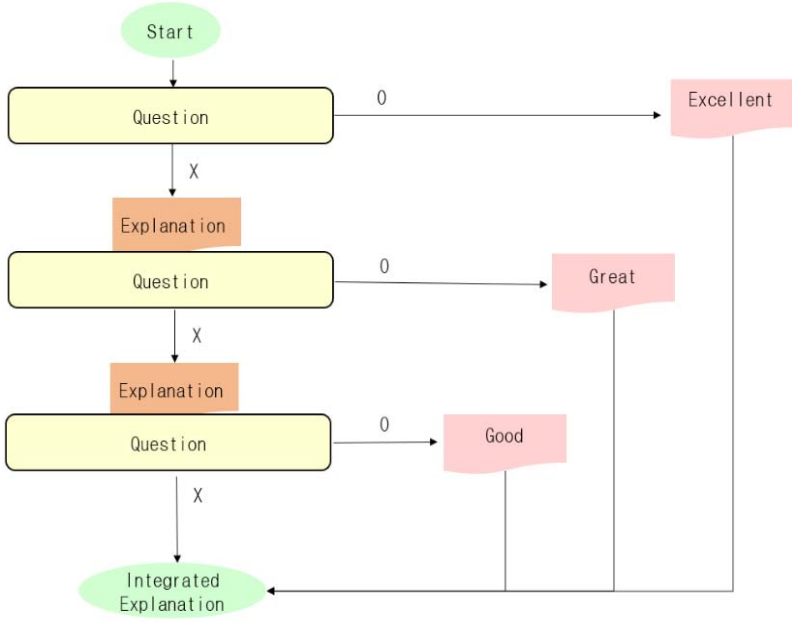

Figure 4. The dialog flow of the system-driven chatbot

by computers, that is, expressing words as vectors through artificial neural network learning. BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained language model (2018, by Google) that uses Contextual Embedding that solves the problem of distinguishing contexts that previous word embedding methods like Word2Vec or GloVe could not. koBERT is the Korean version of BERT. BERT is implemented using Transformer, whereby some of the input is randomly masked and predicted for pre-training (Ahn & Yoo, 2023)

In this study, the meaning of normative grammar follows the broad view of Min (2003), as mentioned in Chapter 2; however, since the target learner follows the 2015 revised Korean language curriculum, the scope of the development area was limited to orthography, pronunciation, and fundamental phonology. As mentioned by Purpura (2004), Lee (2008), Ko (2010), Yi (2013), Kim et al. (2007), and Nam (2007), grammatical ability is divided into knowledge and utilization abilities. In this study, these two aspects were divided into the terms "comprehension" and "application" used by Kim et al. (2007) and Nam (2007). Therefore, the items were divided into the comprehension and application categories. The number of questions per area was 100, and comprehension items (2n–1) and application items (2n) alternated. The comprehension items were organized to move from general to detailed rules, and the application items were presented by changing the sentence cases for the same detailed rule. All questions were limited to two options,[3] and instructions such as "Please choose the appropriate one between 1 and 2 and enter only the numbers" were presented together. The user interface of the system-driven chatbot is shown in Figure 5.

3 This is to reduce the probability that chatbots will perceive the right answer incorrectly. Chatbots are affected by the length of the answer. For example, the similarity between *morpheme* and *morphine* can be recognized as being higher than that between *morpheme* and *word form*. Therefore, if students answer freely and the length of the answer is too short, a problem may arise in recognizing answers.

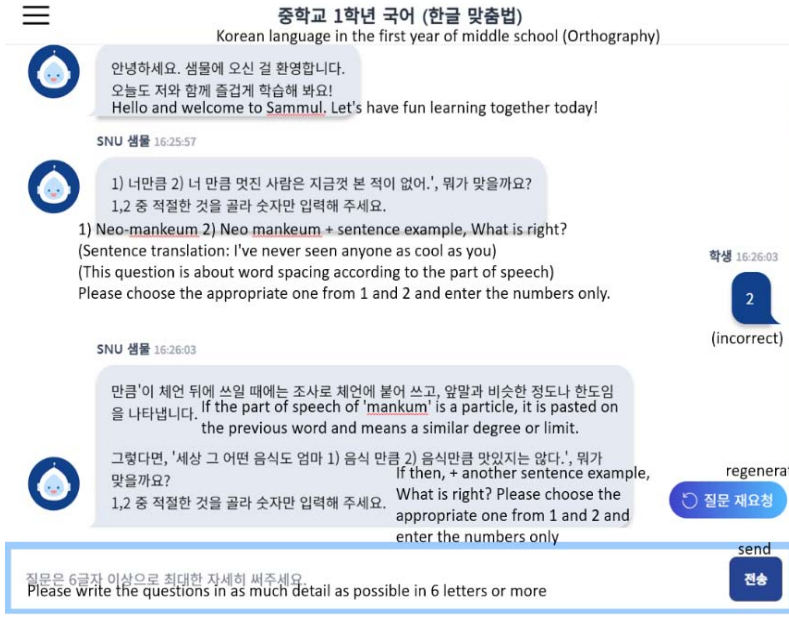

Figure 5. The user interface of the system-driven Chatbot

2) Questionnaires

A questionnaire was developed to investigate chatbot usage patterns and perceptions. The usage pattern was multiple-choice and the items asked about frequency and style. The usefulness evaluation asked about three affordances according to Huang et al.'s (2022) criteria, and the usefulness evaluation and attitude toward the chatbots were elicited on a 5-point Likert scale.[4] The respondent's preference for an overall impression of chatbots was elicited with open-ended questions, and the reliability of the items was confirmed using Cronbach's alpha. Because this value of all individual items was greater than 0.7, all the items were used in the analysis. Table 3 presents the detailed composition of the questionnaire and the overall Cronbach's

4 In order to reduce the damage even if an inconsistent response to one item is made, it is recommended to create more than 4 items per sub-area and use multi-item scales (Dörnyei, 2003, p. 34), so the minimum number of items was set to 4.

alpha values for the categories developed using the Likert scale.

Table 3. Questionnaire Details

Category	Subcategory		Item type	*N*	Cronbach's α
Usage patterns	Frequency and style		Multiple choice	8	
Perception	Usefulness	Technical affordances	Likert (5-point scale)	4	0.864
		Pedagogical affordances	Likert (5-point scale)	6	0.963
		Social affordances	Likert (5-point scale)	4	0.907
	Attitude		Likert (5-point scale)	6	0.972
	Preferences		Open questions	4	
	Overall impression		Open questions	2	
Total				34	

4. Analysis

When analyzing the survey results, multiple responses were allowed for multiple-choice and open-ended question items; therefore, only descriptive statistics were presented. The Likert-scale questions were examined for differences between detailed items using repeated-measures analysis of variance (ANOVA). Normality was confirmed by the skewness and kurtosis values,[5] which were satisfactory. Homoscedasticity was confirmed using the Levene test of the Lawstat

5 According to George and Mallery (2010), values of skewness and kurtosis between −2 and +2 are considered acceptable for the applicability of a normal univariate distribution (Simon, 2018); Aminu and Shariff (2014), who tried rules of thumb to check normality, followed Kline (2011) that an absolute value of skewness greater than 3 and a kurtosis value greater than 10 may indicate a problem and values above 20 may indicate a more serious problem.

package in R and all results were satisfactory. For sphericity, automatic calibration values were reported using the analysis of variance test of the rstatix package in R, retaining a confidence level of 95% and setting the significance level (α) at .05.

IV. Results

1. Usage patterns of chatbots

Regarding chatbot usage patterns, chatbots were mostly used once or twice a week for less than 10 minutes. The results of the frequency analysis are presented in Table 4.[6]

Table 4. The frequency analysis results (*N* = 51)

F2: Weekly Use	Under 1	1-2	3-4	5-6	Over 7
	16	28	6	1	0
F3: Duration (min)	**Under 10**	**10-20**	**20-30**	**30-40**	**Over 40**
	36	13	2	0	0

Typically, students use chatbots at home or after class school to ask questions in the orthographic area about "what." The most prevalent reason for its use is for reviews. Table 5 presents the results of the style analyses.

6 F1 (daily use) questions were excluded from the analysis because students did not understand the questions well. Additionally, the total number of responses after excluding missing values and unfaithful responses was 51.

Table 5. The style analysis results (Allow multiple responses)

F4: Timing	Before class	In class	After class	Etc.
	8	19	22	5
F5: Location	Home	School	Private institute	Etc.
	26	25	2	5
F6: Question area	Orthography	Pronunciation	Phonology	Etc.
	28	22	17	6
F7: Question type	What	How	Why	Etc.
	24	12	23	3

F8: Reason	Motivation	Prior learning	Review	Information search	Etc.
	9	5	24	17	5

2. Perceptions toward chatbots

1) Usefulness

Regarding the technological affordances of chatbots, students perceived that they were highly accessible and not personalized. Descriptive statistics and segment bar plots of the analysis results are presented in Table 6 and Figure 6. The results of checking the difference in scores by survey item found no significant difference between the four questionnaire items ($F_{(3,196)}$=2.527, p=0.07); however, a significant difference between T2 and T4 was observed under the Bonferroni post-test[7] (p<.05).

7 Bonferroni is one of available multiple comparison tests after one-way ANOVA. It controls overall error rate by setting the error rate for each test to the experimentwise error rate divided by the total number of test, and thus, the observed significance level is adjusted for the fact that multiple comparisons are being made (IBM, 2023)

Table 6. Descriptive statistics of technical affordances (*N* = 51)

Item	T1: Timeliness	T2: Accessibility	T3: Interface	T4: Personalization
M(SD)	3.78(1.03)	3.94(1.01)	3.8(0.98)	3.61(1.13)
skew/kurt	–0.77/0.25	–0.92/0.69	–0.74/0.55	–0.42/–0.73

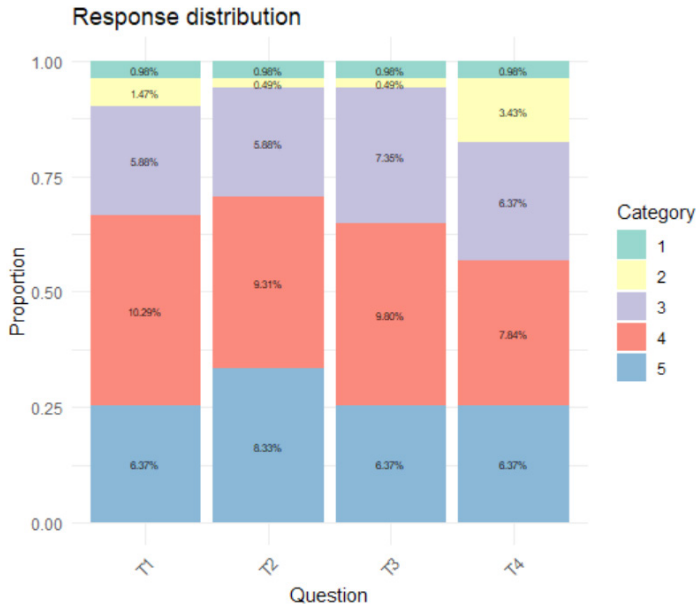

Figure 6. Segment bar plots of technical affordances

Regarding the pedagogical affordance of chatbots, students perceived that although they were well-equipped to improve their knowledge, they were not good at recommending sites or links necessary for studying. Descriptive statistics and segment bar plots of the analysis results are shown in Table 7 and Figure 7. Checking the difference in scores by survey item showed a significant difference between the six questionnaire items ($F_{(5,294)}$= 15.429, p<.05, η^2=0.108), and a significant difference between P6 and other items was observed as a result of the Bonferroni post-test (all p<.005).

Table 7. Descriptive statistics of pedagogical affordances (*N* = 51)

Item	P1: Knowledge	P2: Strategy	P3: Authenticity	P4: Trans-mission	P5: Helpline	P6: Recommendation
M(SD)	3.9(0.94)	3.8(0.98)	3.88(097)	3.8(0.92)	3.69(1.05)	2.92(1.06)
skew/kurt	–0.51/–0.08	–0.61/0.39	–0.54/–0.19	–0.53/0.16	–0.39/–0.69	0.05/–0.42

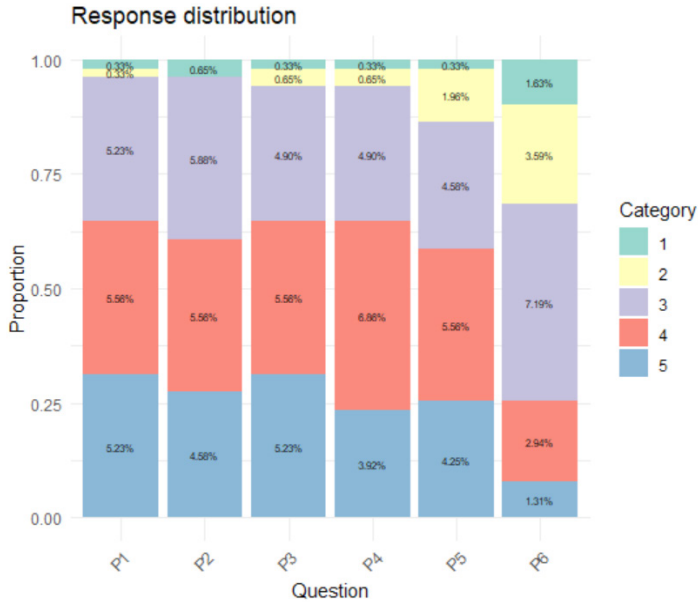

Figure 7. Segment bar plots of pedagogical affordances

Regarding the chatbot's social affordance, students gave a very high score for not being nervous when asking a chatbot (i.e., ≥ 4), but the lowest score was given for being open enough to tell personal stories. The descriptive statistics and segment bar plots of the analysis results are shown in Table 8 and Figure 8. The results of checking the difference in scores by survey item showed a significant difference between the four questionnaire items ($F_{(3,196)}$=16.962, p<.05, η^2=0.119), and a significant difference between S3 and the other items (all p<.005) and between S1 and S4 (p<.05) was observed in the Bonferroni post-test.

Table 8. Descriptive statistics of social affordances (*N* = 51)

Item	S1: Familiarity	S2: Generosity	S3: Openness	S4: No anxiety
M(SD)	3.65(1.07)	3.8(0.98)	3(1.11)	4.02(1.01)
skew/kurt	–0.33/–0.48	–0.49/–0.27	–0.09/–0.83	–0.72/–0.17

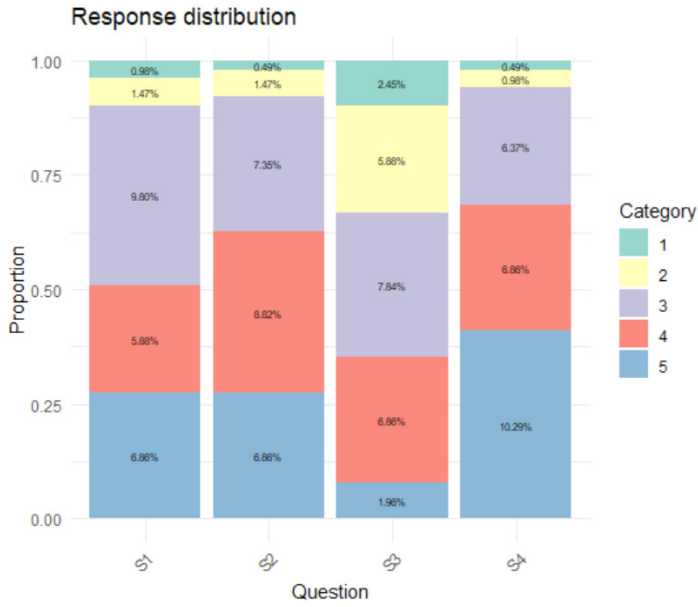

Figure 8. Segment bar plots of social affordances

2) Attitudes

Regarding attitudes toward chatbots, students scored high on the belief that using chatbots would improve their grammatical knowledge but relatively low on fun and willingness to continue their use. Descriptive statistics and segment bar plots of the analysis results are presented in Table 9 and Figure 9, respectively. Checking the differences in scores by survey item found no significant difference between the six questionnaire items ($F_{(5,294)}$=5.648, p<.05, η^2=0.04), and a significant difference between the top three items, AT6, AT3, and AT1, and the bottom three items, AT5, AT4, and AT2, was observed as a

result of the Bonferroni post-test (all p<.05).

Table 9. Descriptive statistics of attitudes (N = 51)

Item	AT1: Novelty	AT2: Fun	AT3: Ease	AT4: Will	AT5: Recommendation	AT6: Belief
M(SD)	3.76(1.09)	3.37(1.08)	3.84(0.97)	3.49(1.07)	3.61(1.04)	3.98(0.91)
skew/kurt	–0.72/ 0.16	–0.10/ –0.70	–0.73/ 0.68	–0.56/ –0.20	–0.33/ –0.31	–0.60/ 0.24

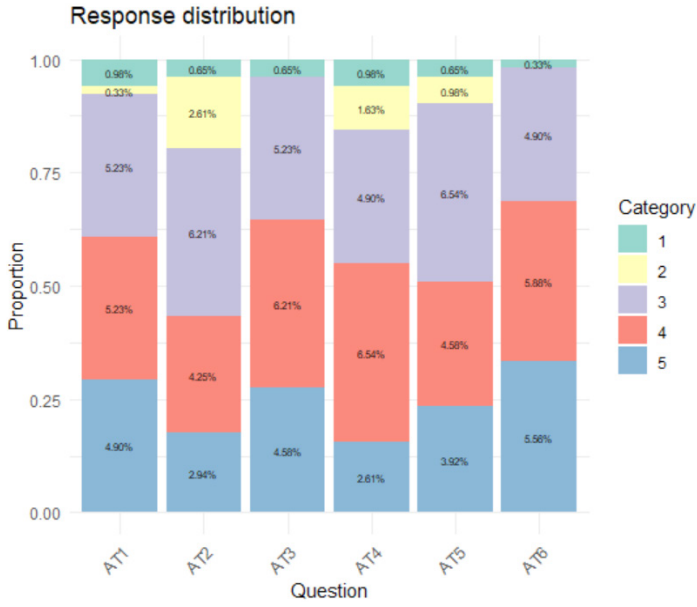

Figure 9. Segment bar plots of attitudes

3) Preferences

Regarding the preference for chatbots, 34% preferred system-driven chatbots and 55% Q&A chatbots, the former because Q&A chatbots immediately answered questions and were free and convenient. Conversely, those preferring system-driven chatbots reported that the Q&A chatbot's answers were inaccurate, or they did not want to ask

questions first. Some participants responded that the chatbot's questions were like a game that motivated them to study further. Regarding questioning areas, the preference for orthography was the highest at 49%, followed by phonology at 24% and pronunciation at 16%. Among the reasons for the preferences in the three areas, a common reason given was that the area was challenging. Orthography is useful in real life, and the need for knowledge checks has been consistently noted. Additionally, the opinions were that it was fun or systematic; this opinion was the same in the phonological area, in particular, that it was effective for the review of pronunciation. The preferences for chatbots and question areas are illustrated in Figures 10 and 11, respectively.

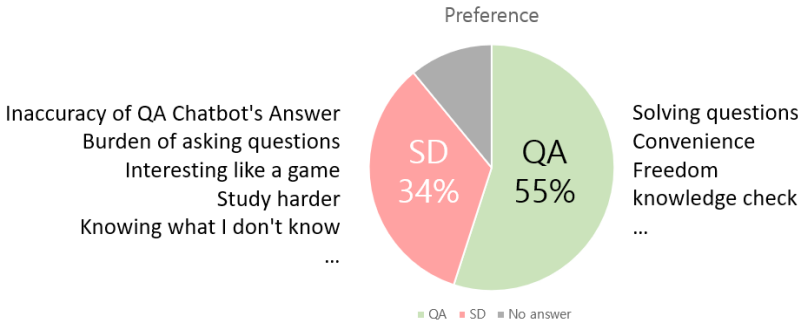

Figure 10. Chatbot preferences

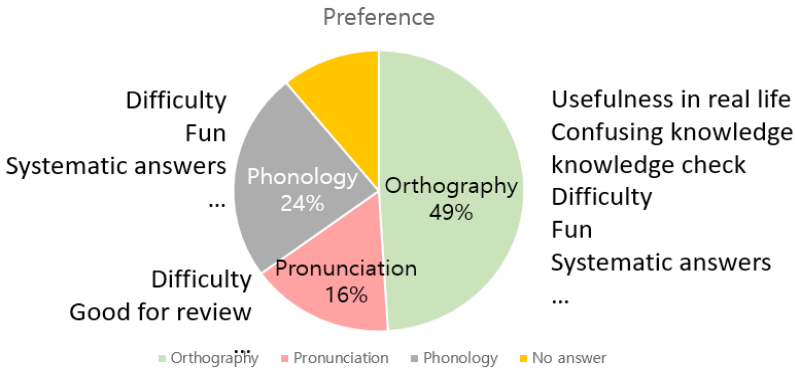

Figure 11. Question area preferences (multiple responses allowed)

4) Overall impressions

The opinions submitted in the survey of the respondents' overall impressions of the last chatbot, divided into positive and negative aspects, are summarized in Figure 12.

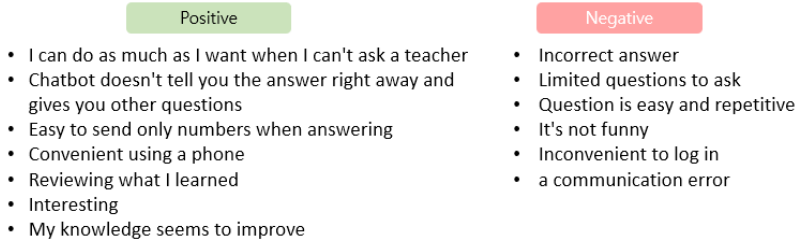

Figure 12. Positive and negative sides of chatbots

V. Discussion and Conclusion

The chatbot developed in this study corresponds to a "prototype for developing chatbots aimed at self-directed learning" by asking questions about a vast range of normative grammar that is challenging to deal with during class, and this paper is part of the process. According to Haristiani's (2019) structural criteria, the developed chatbots were tree-based flow chatbots, which were first asked by students, and system-driven chatbots, which led the questioning. The chatbot survey was conducted using a written questionnaire, and the framework of technical, pedagogical, and social affordances proposed by Huang et al. (2022) was used to evaluate the usefulness of the chatbots.

The survey results have the following implications for future chatbot development, applications, and research. First, database construction and data refinement are required for chatbot development. As this study shows that students most frequently used the chatbot

"for review purposes," it is important to consider this objective when building question-answer pairs in the database. It is also essential to consider the educational nature of chatbots. While many chatbots emphasize brevity and convenience as their primary advantages, this study used multi-turn interactions and provided detailed responses, which contributed to student satisfaction. Therefore, for educational purposes, responses should be detailed to ensure understanding. Second, it is necessary to develop chatbots that are more mobile-friendly. Middle school students use mobile phones more often than computers. The chatbot in this study received "high scores in the accessibility" in the survey, primarily due to its mobile accessibility. Third, there is a need to develop a chatbot with flexible responses, which is classified as artificial intelligence by Haristiani (2019). The chatbot in this study was tree-based and capable of providing fixed responses, which led to "low response accuracy" and "limited questions," with frequent mention of "a lack of openness." This structural issue may lead to a "decrease in fun and willingness to use"; therefore, introducing a different chatbot structure may help address these concerns.

Regarding the implications of chatbot applications, because students primarily use chatbots for review, teachers should design lessons that actively incorporate chatbots into the review process, as by assigning post-class chatbot-related assignments. Second, teachers must provide clear explanations and guidance regarding chatbot use. One of the researchers in this study, a teacher, provided thorough guidance for chatbot usage. This may have led to "strong beliefs in improving their knowledge through chatbots." In the overall impression evaluation, students mentioned the teacher's guidance regarding this belief. Therefore, chatbots can be used effectively only after sufficient instruction and explanations. Third, students may use the chatbot for emotional support along with curricular knowledge. In the survey, students rated the factor of no anxiety highly and mentioned the friendliness of the chatbots. This aligns with the findings of Davis (2022), who demonstrated that chatbots could alleviate tension and

embarrassment.

Finally, there are some implications for chatbot research. First, students scored knowledge enhancement highly in the pedagogical affordances of the survey. However, further research is required to determine whether there are statistically significant differences in the degree of knowledge enhancement. Second, additional research on chatbot preferences is needed. While the study found a strong preference for Q&A chatbots, system-driven chatbots, due to their multi-turn response approach, offer more educational value by allowing for exploration, as mentioned by students in their overall impression evaluation. Therefore, further research is necessary to determine the type of chatbot that can satisfy learners' demands and effectively achieve educational objectives.

This study is significant in that it proposes a new teaching and learning method in normative grammar education. By focusing on the vastness of the normative grammar range and memorization-oriented teaching methods, we explored the possibility of learners' self-directed learning using chatbots. Additionally, while chatbot research has been limited to learners of a specific language and age, this study is innovative in developing chatbots for the language education of native middle school speakers.

As revealed by the survey results, a limitation of this study was its use of a traditional chatbot that provided fixed answers from a limited database, resulting in interaction and educational problems. In particular, LLMs that can already be flexibly answered, such as ChatGPT, BingChat, Bard, and Claud, have been developed; therefore, the effectiveness of the chatbot in this study may be questioned. However, the chatbot of this study differs in that it was based on practical and educational data, as educational experts wrote answers based on actual queries raised by middle school learners who learned normative grammar, unlike popular LLM. In addition, as ChatGPT was not possible to answer the questions used in this study accurately, such as examples and phonological analysis of Korean sentences, chatbots

developed for special purposes may be more useful than popular LLMs in certain areas with very narrow domains. A comparison between the existing LLM and the developed chatbot will be the subject of future research. If more student and teacher question-and-answer data are accumulated and chatbot utilization methods improve, students can expect to develop a reliable chatbot that can be used after school in actual middle schools.

* Submitted 2023.10.30.
First revision recieved 2023.11.21.
Accepted 2023.12.8.

REFERENCES

Ahn, S. J., & Yoo, W. J. (2023). *Introduction to natural language processing using deep learning*. WikiDocs. https://wikidocs.net/book/2155

Aminu, I. M., & Shariff, M. N. M. (2014). Strategic orientation, access to finance, business environment and SMEs performance in Nigeria: data screening and preliminary analysis. *European Journal of Business and Management, 6* (35), 124-132.

Ayedoun, E., Hayashi, Y., & Seta, K. (2015). A conversational agent to encourage willingness to communicate in the context of English as a foreign language. *Procedia Computer Science, 60*, 1433-1442.

Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: changing user needs and motivations. *Interactions, 25*(5), 38-43.

Chen, H. L., Widarso, G. V, & Sutrisno, H. (2020). A chatbot for learning Chinese: learning achievement and technology acceptance. *Journal of Educational Computing Research, 58*(6), 1161-1189.

Davis, G. M. (2022). *Humanlike conversational artificial intelligence agents for foreign or second language learning: User perceptions, expectations, and interactions with agents*. Stanford University.

Dörnyei, Z. (2002). *Questionnaires in second language research: construction, administration, and processing*. Mahwah, N. J. : Lawrence Erlbaum Associates.

Fidan, M., & Gencel, N. (2022). Supporting the instructional videos with chatbot and peer feedback mechanisms in online learning: The effects on learning performance and intrinsic motivation. *Journal of Educational Computing Research, 60*(7), 1716-1741.

Følstad, A., & Brandtzaeg, P. B. (2020). Users' experiences with chatbots: findings from a questionnaire study. *Quality and User Experience, 5*(1), 3.

George, D., & Mallery, M. (2010). *SPSS for windows step by step: A simple guide and reference*. 17.0 update (10a ed.). Boston: Pearson.

Goda, Y., Yamada, M., Matsukawa, H., Hata, K., & Yasunami, S. (2014). Conversation with a chatbot before an online EFL group discussion and the effects on critical thinking. *The Journal of Information and Systems in Education, 13*(1), 1-7.

Haristiani, N. (2019). Artificial Intelligence (AI) chatbot as language learning medium: An inquiry. *In Journal of Physics: Conference Series, 1387*(1), 1-6.

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning: Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning, 38*(1), 237-257.

IBM. (2023). One-way ANOVA post hoc test. *In IBM SPSS Statistics subscription documentation*. Retrieved December 08, 2023, from https://www.ibm.com/docs/en/spss-statistics/saas?topic=anova-one-way-post-hoc-tests

Kharis, M., Schon, S., Hidayat, E., Ardiansyah, R., & Ebner, M. (2022). Mobile gramabot: Development of a chatbot app for interactive german grammar learning. *International Journal of Emerging Technologies in Learning, 17*(14), 52-63.

Kim, E. S., Nam, K. Y., Kim, H. J., & Park, J. H. (2007). On the phonological knowledge of Korean grammar learners. *The Journal of Linguistics Science, 42*, 1-25.

Kim, J., Oh, S., Kwon, O. W., & Kim, H. (2019). Multi-turn chatbot based on query-context attentions and dual wasserstein generative adversarial networks. *Applied Sciences, 9*(18), 3908.

Kim, N. Y. (2016). Effects of voice chat on EFL learners' speaking ability according to proficiency levels. *Multimedia-Assisted Language Learning, 19* (4), 63-88.

Kim, N. Y. (2018a). Chatbots and Korean EFL students' English vocabulary learning. *Journal of Digital Convergence, 16*(2), 1-7.

Kim, N. Y. (2018b). A study on chatbots for developing Korean college students' English listening and reading skills. *Journal of Digital Convergence, 16*(8), 19-26.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (5th ed., pp. 3-427). New York: The Guilford Press.

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(3), 1-14.

Koo, B. K. (2010). A study on grammatical competence and the direction of grammar assessment. *Korean Language Education Research, 37*(0), 185-218.

Lee, J., Kim, H., Shin, D., & Yang, H. (2019). A study on the conversational interaction with chatbot for foreign language learning. *Multimedia-Assisted Language Learning, 22*(1), 132-153.

Lee, K. K. (2008). *The theory of school grammar education*. Korea University Ethnic Culture Research Institute.

Min, H. S. (2003). The direction of the language norm policy. *Journal of Korean Language Education, 12*(0), 7-71.

Min, Y. J., Ahn, J. K., & Kim, S. Y. (2020). Structural relationship analysis between the intention to use educational chatbots and influential factors based on technology acceptance model. *Journal of Korean Association for Educational Information and Media, 26*(4), 799-825.

Nam, G. Y. (2014). A study of relationship between Korean normative grammar education and grammar education: Based on analysis of grammar curriculum and textbook. *The Korean Language and Literature, 62*(0), 1-44.

Nam, K. Y. (2007). The systematization of "structure of knowledge" in grammar education, *Korean Language Education. 123*, 341-374.

National Institute of Korean Language, https://kornorms.korean.go.kr/main/main.do

Paik, Y. K. (2021). A study on developing chatbot for teaching Korean pragmatics: Focusing on the refusal speech act. *Bilingual Research, 84*, 115-141.

Park, C. (2023, June 5). 1st place Claude put ChatGPT, Bingchat, and Bard to shame. *AI Times*. https://www.aitimes.com/news/articleView.html?idxno=151525

Park, J. A., & Lee, H. (2021). AI chatbot builders utilizaion plan to develop AI chatbots for Korean language education. *Teaching Korean as a Foreign Language, 63*, 51-91.

Polanyi, M. (1962), *Personal Knowledge: Towards a Post-critical Philosophy*, Chicago: University of Chicago Press.

Purpura, J. E. (2004). *Assessing grammar*, Cambridge University Press.

Simon, S. (2018). Testing normality including skewness and kurtosis. *In MRC CBU Wiki*. University of Cambridge. Retrieved from December 08, 2023, from https://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/Simon

Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the Facebook Messenger. *Computers & Education, 151*, 1-11.

Suh, J. H. (2021). A study on the design method of Korean speaking practice chatbot for beginners based on scenarios. *Culture and Convergence, 43*, 53-76.

Wang, Y. F., Petrina, S., & Feng, F. (2017). VILLAGE—Virtual immersive language learning and gaming environment: Immersion and presence. *British Journal of Educational Technology, 48*(2), 431-450.

Yi, K. Y. (2013). A study for grammar assessment for grammatical judgement: With grammatical judgment. *Journal of Korean Language Education. 32*(0), 105-132.

Yoo, J. Y., & Yoo, H. S. (2021). The effects of smart-learning in Korean language education through chatbots: A focus on learner perception. *Multimedia-Assisted Language Learning, 24*(2), 82-105.

Yu, H. R. (2017). A Study on 'grammar' and 'activity' as contents of grammar education. *Eo Mun Lon Chong (Korean Language and Literature)*, 72(0), 81-113.

ABSTRACT

Learners' Experiences with Artificial Intelligence (AI) Chatbot for Korean Grammar Education

: A Focus on Usage Patterns and Perception

Song, Changkyung · Seo, Heeju · Kim, Hojung

This study investigated the usage patterns and perceptions of normative grammar education in the Seoul National University, ETRI/TutorousLabs, and the Timbel Collaboration Project. Two types of chatbots were developed: a Q&A chatbot with students initiating questions and a system-driven chatbot with the chatbot initiating questions. Convenience sampling was employed to recruit 62 middle school students for the survey, which covered aspects such as frequency, style, usefulness, technical affordances, pedagogical affordances, social affordances, attitude, preference, and overall impression. Usefulness and attitude were scored on a 5-point Likert scale, whereas other items involved multiple-choice and open-ended questions. The findings were as follows: chatbots were used 1-2 times per week for less than 10 minutes. Several questions were of the "what" type, focusing on orthography areas at home or school for review purposes. Students rated the accessibility and knowledge improvement highly, with no reported anxiety about chatbot use. However, the ratings were relatively low for personalization, recommendations, and openness. Regarding attitude, there was a strong belief in knowledge improvement and a weak inclination to continue use. Regarding preferences, orthography and the Q&A chatbots were dominant. Finally, this study provides insights into the development, application, and research on chatbots in educational contexts.

KEYWORDS Chatbot, Q&A chatbot, System-driven chatbot, Normative grammar education, Usage pattern, Perception