Exploring User-Friendly AI Tools and Their Emotional Impact on Japanese ESL Learners

PARKIN Douglas Robert*

(Abstract)

Most educators view the Japanese ESL education system as needing significant changes, particularly in terms of communication, especially oral communication. This mixed-methods repeated-measures study aimed to identify the most user-friendly AI tools that Japanese ESL learners perceived as effective for improving their oral and written communication skills while promoting long-term intrinsic motivation. The study was conducted with 114 students from six classes at two Japanese universities, in the faculties of education, engineering, and medicine. It was conducted over two months, utilizing two sub-studies: one a four-week oral communication study that evaluated ChatGPT, Pi, Gemini, and Bing, and the other a four-week written communication study that evaluated Grammarly, Wordtune, ProWritingAid, and QuillBot. The oral and written communication studies followed parallel methods, with students interacting with AI apps on assigned topics, receiving skill assessments, and completing Google Forms questionnaires about their experiences. However, because questionnaire responses were anonymous and lacked a stable participant identifier, tool comparisons for the Yes/No outcomes are presented descriptively (percentages by tool), and differences are interpreted as trends. 13 interviews were conducted at the end of the eight weeks to gain further insights. The findings show that students favoured ChatGPT and Grammarly for oral and written communication, viewed feedback as consistent/helpful, and valued the emotional support provided by the AI tools. They also appreciated the apps' user-friendliness, which offered immediate, positive, and tailored feedback that supported the development of oral and written communication skills, fostered autonomy and kindness, and promoted long-term intrinsic motivation. This study provides a deeper understanding of how AI can be used to enhance ESL proficiency while supporting learners cognitively and emotionally throughout the process.

Keywords

User-friendly, social-emotional, AI, ESL, oral and written communication, skills development, intrinsic motivation

Introduction

Despite years of efforts to improve ESL education in Japan, the country ranked 92nd globally for English proficiency in 2024, its worst-ever ranking (EF Epi, 2024). The struggles of Japanese students, particularly with oral English, are highlighted by a national test where third-year junior high students averaged only 30.8% in speaking (Long & Watanabe, 2021).

* The Graduate School of East Asian Studies, Yamaguchi University

Journal of East Asian Studies, No.24, 2026.3. (pp.173-195)

To address this, educators should move beyond traditional grammar-translation methods and provide authentic communication opportunities essential to ESL education in Japan (Abe, 2013; Morita, 2017). AI is widely regarded as a viable means of providing authentic communication opportunities, owing to its recent popularity and growing research evidence supporting these purposes (Du & Daniel, 2024; Wang, 2025). Studies show that AI-supported ESL oral practice can offer authentic, personalized interaction within calm, supportive environments that some learners prefer to human-led instruction (Almelhes, 2023; Wang, 2025). Research shows that in written communication, AI offers learners human-like suggestions and provides assessments that can equal or even surpass human evaluators (Alharbi, 2023; Mizumoto et al., 2024). While AI shows promise in ESL education, research highlights the need for more rigorous studies, particularly by ESL specialists in collaboration with experts in other fields (Sharadgah & Sa'di, 2022; Son et al., 2023).

In addition, before educators adopt this new technology simply because it is popular, we must first consider how to preserve the human factor in the educational process. UNESCO's 2021 Recommendation on the Ethics of Artificial Intelligence states that AI systems should never replace human decision-making autonomy and must always enhance quality of life, while empowering teachers and students by supporting, rather than diminishing, the learning process (UNESCO, 2022). To support these recommendations and ensure this study was conducted in the best interests of ESL educators and learners, the AI Directive was developed and applied to this research. The AI Directive is a philosophical and pedagogical approach that stipulates AI should never be used to create or replace humans and their contributions, which can and should be made independently. Instead, AI should only supplement and enhance learning while always maintaining learner autonomy! In line with this directive, the purpose of this study was to identify the most user-friendly AI tools that Japanese ESL learners perceive as effective in enhancing their oral and written communication skills while fostering long-term intrinsic motivation. Supporting this purpose are three research questions:

1. What free, user-friendly, and easily accessible AI tools are available to ESL learners to improve their oral and written communication skills on a sustained basis?
2. What AI tools do ESL learners perceive as effective for improving their oral and written communication skills, and what factors influence these perceptions?
3. What AI tools intrinsically motivate ESL learners, align with their emotional and psychological needs, and effectively promote sustainable long-term commitment to improving oral and written communication skills?

Literature Review

Need for Changes to the Japanese ESL Educational System

Most English educators in Japan and worldwide, familiar with the Japanese ESL system, agree that drastic changes are needed, as evidenced by Japan's record-low 92nd place among 116 countries in the 2024 global English proficiency ranking (EF EPI, 2024). In Japan, ESL skills are vital but challenging to maintain or improve, as English is rarely used in daily life. A study found that Japanese EFL university students in Malaysia, despite formal learning, still experienced higher oral communication apprehension than other students (Iwamoto, 2016; Jalleh et al., 2021). Many reforms in elementary English education stem from the belief that Japan's low English proficiency could lead to its political and economic decline (Terasawa, 2022). Despite numerous initiatives by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) over the past decade to improve ESL education, studies indicate a persistent gap between policy objectives and classroom realities, underscoring the need for curriculum reform, professional development, and resource reallocation (Smith, 2025). English education in Japanese elementary schools began as a foreign-language activity in 2011 and became a compulsory subject in 2020, with MEXT assigning instruction to many homeroom teachers who themselves struggled with English, resulting in widespread problems (Iwamoto, 2016; Takiguchi & Machida, 2020; Terasawa, 2022).

Japan recognizes that English communication skills are essential in a global society, yet many remain reluctant to speak due to low self-confidence despite years of study (Abe, 2013). This reluctance is reinforced by an exam-focused education system that neglects communication and fosters anxiety, highlighting the need for reform toward communicative methods that build positive attitudes and practical skills (Morita, 2017; Samuell, 2021). Supporting this, studies indicate that students receiving English-medium instruction (EMI) in Japanese high schools perform more successfully at EMI universities than peers taught through traditional methods (Aizawa & Rose, 2020). Students who experience using English abroad are often viewed as cool in Japan, but this can also lead to anxiety and isolation due to the attention they receive (Sakamoto & Furukawa, 2022). This highlights the need for an accepting English learning system that prioritizes self-improvement over comparative positioning. Mainstream perceptions and advertising strongly shape how young learners perceive English communication in Japan. Many English conversation schools (eikaiwa) use psychologically driven marketing that positions them as solutions to employer pressure to improve English skills, often sending mixed messages to learners (Nuske, 2019).

AI as a Viable Option to Support ESL Education

To address deficiencies in Japanese ESL education, including exam-driven and grammar-focused instruction, research suggests that AI can provide personalized support to build

fluency, foster creativity, and reduce speaking anxiety (Busso & Sanchez, 2024). Two 2022 meta-analyses concluded that AI integration in language learning significantly enhances oral communication and writing quality, improves affective factors like attitude and interest, and provides practical assessment and translation tools, with one study emphasizing human-AI collaboration as key to effective learning (Yang & Kyun, 2022; Sharadgah & Sa'di, 2022). However, they also found that more rigorous studies were needed for AI in ESL education. Another study found AI chatbots outperformed traditional EFL teaching methods by providing immediate feedback, improving vocabulary, and triggering positive emotions, though interest declined as the novelty faded (Wu & Li, 2024).

AI shows great promise for developing ESL oral skills, as speech recognition can detect pronunciation errors and track performance, often boosting learner confidence more than human teachers (Almelhes, 2023). While some doubt that AI can provide authentic English communication, chatbots, unlike text-only platforms such as Wikipedia, offer oral interactions with unlimited responses and personalized feedback on language and emotion (Du & Daniel, 2024). As AI models like ChatGPT-4 evolve, ESL learners benefit from authentic, personalized conversations that foster skill development in low-stress environments, creating a sense of calm and support (Wang, 2025). Research shows that ChatGPT provides more engaging dialogue materials than traditional textbooks, boosting learner confidence, improving oral skills, and supporting real-world practice, such as delivering a TED Talk (Lo et al., 2024).

AI significantly enhances ESL written communication, as research shows AI writing tools provide indispensable, human-like suggestions for improvement (Alharbi, 2023). While AI can enhance writing, teacher support remains crucial for processing corrections and fostering reflection. A key strength of AI is its accurate writing assessment, with ChatGPT matching human evaluators, and correlating 94% with humans in error detection (Mizumoto et al., 2024; Pfau et al., 2023).

Social Emotional Learning

While early AI-in-ESL studies initially emphasized enjoyment and grade gains, recent research has started to highlight its role in supporting social-emotional learning (SEL). SEL can be defined as the ability to manage emotions, apply skills and knowledge to develop a healthy identity, and achieve cognitive goals while maintaining positive relationships with others (CASEL, 2023; UNESCO, 2024). Integrating SEL into ESL education has been found to enhance emotional well-being by reducing anxiety and boosting engagement, while fostering resilience, communication skills, and a growth mindset (Wicaksono & Saraswati, 2024). Studies show that AI in education enhances students' perspective-taking, problem-solving, self-confidence, and introspective thinking; however, caution is needed to prevent overreliance, which may undermine self-trust and create dependence on AI tools (Capinding & Dumayas, 2024; Chauncey & McKenna, 2023). Studies using AI in ESL lessons have found that it improves English skills

and emotional well-being and may even support basic psychological needs through apps like ChatGPT (Chen et al., 2024; Liu & Reinders, 2025). By creating personalized, emotionally supportive environments, an AI-based SEL framework was shown to enhance linguistic skills, emotional intelligence, resilience, and social abilities among university ESL learners in China (Zong & Yang, 2025).

Although anxiety is common in ESL classrooms, AI can help reduce it through judgment-free interactions, personalization, and real-time feedback, thereby building confidence, resilience, a sense of belonging, and intrinsic motivation (Ahmed et al., 2025; Wu, 2024; Zong & Yang, 2025). An EFL study with Chinese university students found that AI fostered hope, pride, and enjoyment, but these novel emotions also triggered stress, fear, and anxiety (Guo & Wang, 2024). University studies in Hong Kong and Japan found that AI in ESL classes boosted enthusiasm, risk-taking, and reassurance, but also cautioned that over-comparison with AI abilities could undermine learners' confidence, underscoring the need for supportive teacher connections (Kohnke & Moorhouse, 2025; Yamaoka, 2024).

Methodology

Overview and Participants

The study was conducted over two months using a mixed-methods repeated-measures design with two sub-studies: an Oral Communication Study and a Written Communication Study. One hundred fourteen first- to third-year students from six classes at two universities in the faculties of education, engineering, and medicine, participated. Before starting, students provided written consent after being fully informed in English and Japanese about the research, including anonymity and the right to withdraw without penalty.

Oral and Written Sub-Studies

Both sub-studies had students evaluate four different AI tools, one per week, over eight weeks in a fixed weekly sequence. All AI tools were accessed via free web interfaces, using default settings. The oral study evaluated ChatGPT-4o (web interface; accessed Oct-Nov 2024; default settings), Pi (Inflection-2.5; web interface; accessed Oct-Nov 2024; default settings), Gemini (1.x; web interface; accessed Oct-Nov 2024; default settings), and Microsoft Copilot (formerly Bing Chat; web interface; accessed Oct-Nov 2024; default settings). The written study evaluated Grammarly (GO; web interface; accessed Nov-Dec 2024; default settings), Wordtune (AI21; web interface; accessed Nov-Dec 2024; default settings), ProWritingAid (Generative AI; web interface; accessed Nov-Dec 2024; default settings), and QuillBot (Proprietary AI; web interface; accessed Nov-Dec 2024; default settings). The tools were selected because they were freely available, easy to use, and unlikely to impose financial or cognitive burdens on teachers or students, while also supporting skill development and emotional well-being.

Each week, 30 minutes of class time were allocated to introducing the AI tools and topic (5 minutes), students discussing approaches (3-5 minutes), app-based practice and self-testing (15 minutes), and completing a Google Forms questionnaire (5 minutes; see Appendices B and C). To ensure reproducibility and rigour, students were given AI-generated junior-high-level topics on themes such as friends, weather, and travel.

In addition, to support proper learning during the study, the theories underlying the AI Directive were explained to students, reminding them that they should always be in control of the learning process by making their own decisions rather than relying on the apps to decide for them.

Prompts and Questionnaires

During the oral communication sub-study, students installed a new AI application on their phones each week and engaged with it using a standardized oral communication prompt (see Appendix A). The prompt was designed to support and assess oral skills in pronunciation, fluency, vocabulary, grammar, and coherence and cohesion, key elements evaluated by standardized tests such as TOEIC, CEFR, and TOEFL. In addition to chatting, homework, and assessments, the prompt requested positive motivation and supportive feedback to address learners' social-emotional needs. Before starting, students pasted the prompt into the app, entered the daily topic, activated voice mode, and said, "Please start." After approximately ten minutes of chatting, they said, "Please judge me now." The app would then evaluate each standardized element out of 100%, give homework, and then provide social-emotional supportive comments. After receiving feedback, students completed a mixed-methods questionnaire with one Likert-scale item on app usability, along with questions on oral skills, enjoyment, willingness to reuse, and usefulness for discussing daily topics (see Appendix B).

During the written communication sub-study, students installed a new app weekly, with laptops recommended for easier writing. They wrote on levelled topics for about eight minutes, used an AI app to correct their work, then submitted the corrected writing to ChatGPT using a prompt (see Appendix A) designed to assess vocabulary, grammar, content, organization, and coherence and cohesion. ChatGPT was used, given its high level of accuracy in assessing writing (Mizumoto et al., 2024). In addition to writing assessments, the prompt also requested that positive motivation and supportive feedback be given to address learners' social-emotional needs. After receiving ChatGPT's response, students completed a mixed-methods questionnaire including reflections, a seven-point Likert usability rating, percentage-based skill evaluations, and questions on enjoyment, value, and willingness to reuse the app (see Appendix C).

Interviews

After eight weeks, 30-minute semi-structured interviews were conducted with 13 volunteers, about two from each class, using a 12-question guide (see Appendix D). The questions

extended those in the questionnaires, allowing participants to specify which apps they preferred and why, for ESL skills development. Interviews took place in the author's office or a quiet classroom, with assurances of confidentiality, voluntary participation, and data security, including pseudonyms. They were recorded via video, audio, and Otter.ai for transcription.

Data Analysis

Interview and Google Forms data were transferred to a secure PC, where Otter.ai transcriptions were verified, cleaned, manually coded, and categorized, while questionnaire data were merged and analyzed. An example of a code created from Sophia when asked which app she enjoyed most is "I want to use Grammarly…my friends often use Grammarly," resulting in the code "Grammarly is most enjoyable to use and highly recommended by friends." To reduce bias and ensure consistency, the author developed and analyzed 220 interview codes over several weeks, repeatedly cross-checking them with field notes in Microsoft OneNote.

Additionally, the findings were compared multiple times with recordings and other relevant data, such as questionnaires. Qualitative and quantitative analyses were performed in RStudio (Posit Software, 2025) using scripts generated by Gemini Pro for thematic visualizations, word clouds, sentiment analyses, descriptive statistics (e.g., mean, median, mode, standard deviation), and internal-consistency estimates (Cronbach's alpha). Given that questionnaire responses were collected without a stable participant identifier, tool-comparison analyses for the Yes/No outcomes (Q3-Q5) are reported descriptively (percentages by tool) rather than using within-subject inferential tests. Results from RStudio analyses of interviews and questionnaire data were cross-referenced with verbatim student quotes, providing richer context and support. Questionnaires, prompts, RStudio scripts, and other data are available on the Open Science Framework (OSF) at https://osf.io/nr8tc/overview?view_only=7fd31d6386a84dc6b62ea6e621035345.

Results and Discussion

Oral Communication Questionnaires

In the four-week Oral Communication Sub-Study, 424 of a possible 456 responses were collected, yielding a response rate of 92.98%. Missing responses were attributed to possible submission errors and absenteeism. Table 1 presents ChatGPT's results as a cross-section of all four apps, displaying results similar to those of the other apps, with average scores ranging from 70-78%. Students' skill levels averaged in the low 70s overall. Cronbach's alpha indicated high internal consistency among the five oral communication subscale scores, as AI-generated evaluation outputs; however, this does not by itself establish that the scores are valid measures of learners' actual proficiency. This consistency was similar across all apps, with ChatGPT ($\alpha = 0.931$) and Gemini ($\alpha = 0.884$). For ease of use, students rated ChatGPT 4.73/7 as slightly less

difficult than the other AI oral apps, which averaged 5/7.

Table 1 Oral Communication Questionnaire Results of ChatGPT's Skills Evaluations

Statistic Used	APP Used	Pronunciation	Fluency	Vocabulary	Grammar	Coherence and Cohesion	Ease of Use 1 Very Easy to 7 Very Difficult
Mean	ChatGPT	78	72	72	70	72	4.73
Median	ChatGPT	80	75	75	70	75	5.00
Mode	ChatGPT	80	75	70	70	80	5.00
Standard Deviation - *SD*	ChatGPT	12.73	12.50	9.92	11.92	13.30	1.57
Cronbach's Alpha - *α*	ChatGPT	0.931	0.931	0.931	0.931	0.931	

Note. This table was created by the author using RStudio to calculate mean, median, mode, and Cronbach's Alpha (Posit Software, 2025).

Table 2 presents the yes/no responses to Questions 3-5 of the Oral Communication Questionnaire (see Appendix B) regarding the usefulness, enjoyment, and desire to reuse. The expected "Yes/No" answers were often written ambiguously, preventing tools such as RStudio from analyzing them properly and necessitating manual calculation. Yes or No responses were coded based on the overall positivity or negativity of each reply. For example, the response "Soso. This app is so difficult" was coded as No, whereas "Today's topics were easy to talk about" was coded as Yes. Unclassifiable responses were treated as indeterminate and excluded from analysis. ChatGPT received the highest percentages of Yes responses for usefulness (86.11%), enjoyment (81.48%), and willingness to reuse (87.27%). In comparison, Pi, Bing, and Gemini received lower percentages of Yes responses for usefulness (65.39%, 61.68%, 55.14%), enjoyment (66.35%, 58.88%, 46.73%), and willingness to reuse (64.42%, 57.01%, 51.43%), respectively. ChatGPT's high Yes response rates may partly reflect popularity bias, whereas technical issues, including non-English interfaces, inadequate analysis, and app instability, likely influenced Gemini's lower ratings. Although participants evaluated multiple tools, responses were collected without a stable participant identifier, so ratings could not be linked across tools at the individual level. Therefore, within-subject (repeated-measures) statistical inference was not possible for the Yes/No outcomes; we reported descriptive percentages by tool and interpreted differences cautiously.

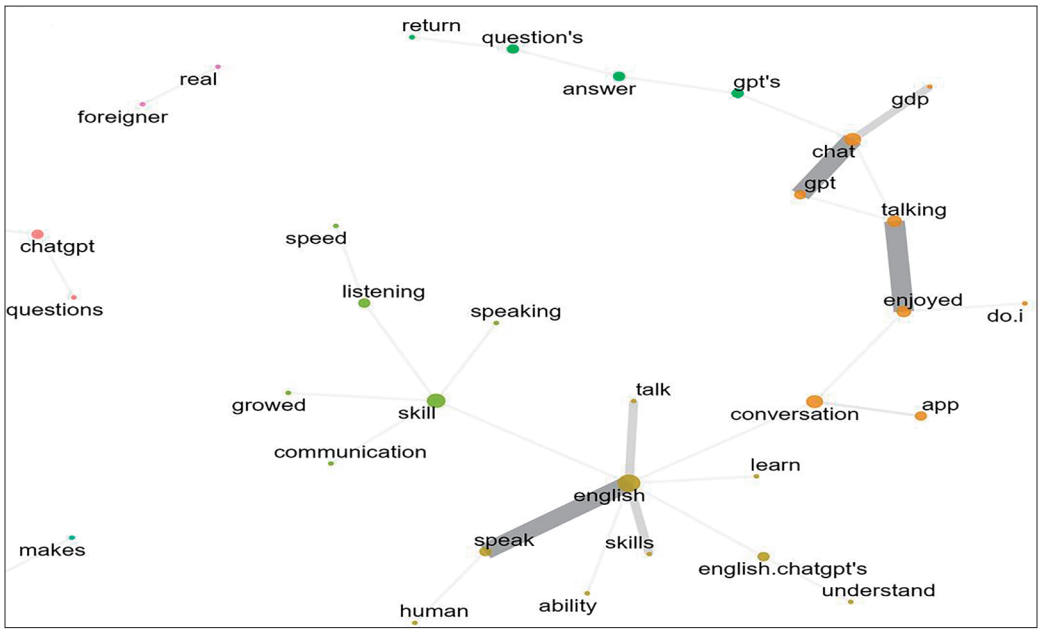
Table 2 Oral Communication Questionnaire Results of all Apps Q3-5 Yes/No Answers

3. Did you enjoy using the app? Please explain.			
ChatGPT	Pi	Bing	Gemini
Yes - 88/108	Yes - 69/104	Yes - 63/107	Yes - 50/107
81.48%	66.35%	58.88%	46.73%
4. Would you use this app again to improve your English or simply to have fun? Please explain.			
ChatGPT	Pi	Bing	Gemini
Yes - 96/110	Yes - 67/104	Yes - 61/107	Yes - 54/105
87.27%	64.42%	57.01%	51.43%
5. Did you find this app useful for discussing today's topic or question? Please explain...			
ChatGPT	Pi	Bing	Gemini
Yes - 93/108	Yes - 68/104	Yes - 66/107	Yes - 59/107
86.11%	65.39%	61.68%	55.14%

Note. This table was created by the author using Word after manually counting questionnaire data.

Figure 1 shows a positive sentiment thematic network created in RStudio for Q3, highlighting words students used to explain why they enjoyed the app. English appeared most frequently with speaking, skills, and talk, while chat, talking, and enjoyed were commonly associated with answer, conversation, and questions. Questionnaire responses support these findings, stating "I was able to talk to the app as if I was talking to a friend" and "Thanks to this activity, I wanna be able to talk English well."

Figure 1 Positive Sentiment Thematic Network - ChatGPT - Q3. Did you enjoy using the app?



Note. This figure was created by the author using RStudio (Posit Software, 2025).

Figure 2 displays a word cloud generated by RStudio, illustrating positive student sentiment towards Q5. Findings suggest that students found the apps useful because they were easy to talk to, felt like friends, and were helpful. Supporting the usefulness of AI, one interview participant, Bill, stated, "Before, I didn't have experience talking with a foreigner… I want to talk with a foreigner." This response highlights the perceived value of AI in providing accessible communication opportunities and motivating continued use. This, of course, is open to other interpretations.

Figure 2 Positive Sentiment Word Cloud - All Apps - Q5. Did you find this app useful for discussing today's topic or question?

Note. Images in this figure were created by the author using RStudio (Posit Software, 2025).

Written Communication Questionnaires

In the four-week Written Communication Sub-Study, 432 of 456 possible responses were collected, yielding a 94.74% response rate. Missing responses were attributed to possible submission errors and absenteeism. Table 3 presents ChatGPT's evaluation of Grammarly-aided writing, with an average of 69.8% across all elements, slightly higher than the 67.47% average for all other apps. Cronbach's alpha indicated high internal consistency among the five AI-generated written communication subscale scores; however, this alone does not establish their validity as measures of learners' actual proficiency. Similar consistency was observed across all apps, including Grammarly ($\alpha = 0.900$) and Wordtune ($\alpha = 0.943$). For ease of use, students rated all apps somewhat difficult, with an average of approximately 4/7.

Table 3 Written Communication Questionnaire Results of ChatGPT's Skills Evaluations

Statistic Used	APP Used	Vocabulary	Grammar	Content	Organization	Coherence and Cohesion	Ease of Use 1 Very Easy to 7 Very Difficult
Mean	Grammarly	69	71	74	68	68	4.02
Median	Grammarly	70	70	75	70	70	4.00
Mode	Grammarly	70	70	80	70	70	4.00
Standard Deviation - *SD*	Grammarly	10.98	11.70	8.65	10.36	8.97	1.68
Cronbach's Alpha - *α*	Grammarly	0.900	0.900	0.900	0.900	0.900	

Note. This table was created by the author using RStudio to calculate mean, median, mode, and Cronbach's Alpha (Posit Software, 2025).

Table 4 presents the yes/no responses to Questions 3-5 of the Written Communication Questionnaire (see Appendix C) regarding enjoyment, willingness to use again, and usefulness for developing and evaluating writing skills. The criteria for determining whether a response was "Yes" or "No" were the same as listed in Table 2. Grammarly received the highest percentages of Yes responses for enjoyment (91.35%), usefulness in developing and evaluating writing skills (91.35%), and willingness to reuse (83.65%). In comparison, ProWritingAid, QuillBot, and Wordtune received lower percentages of Yes responses for enjoyment (88.57%, 85.71%, 83.76%), usefulness (88.57%, 80.95%, 76.07%), and willingness to reuse (73.33%, 79.05%, 80.34%), respectively. Supporting these figures, participants spoke about Grammarly, saying, "It corrects my mistake grammar in an instant", "The app fix all my mistakes", and "Yes, I love this app." Some participants also indicated that they did not appreciate the complexity of certain apps because they offered too many choices and too much information. As in Table 2, given the lack of stable participant identifiers, repeated-measures statistical inference was not possible for the Yes/No outcomes, so we have therefore reported descriptive percentages by tool and interpreted differences cautiously.

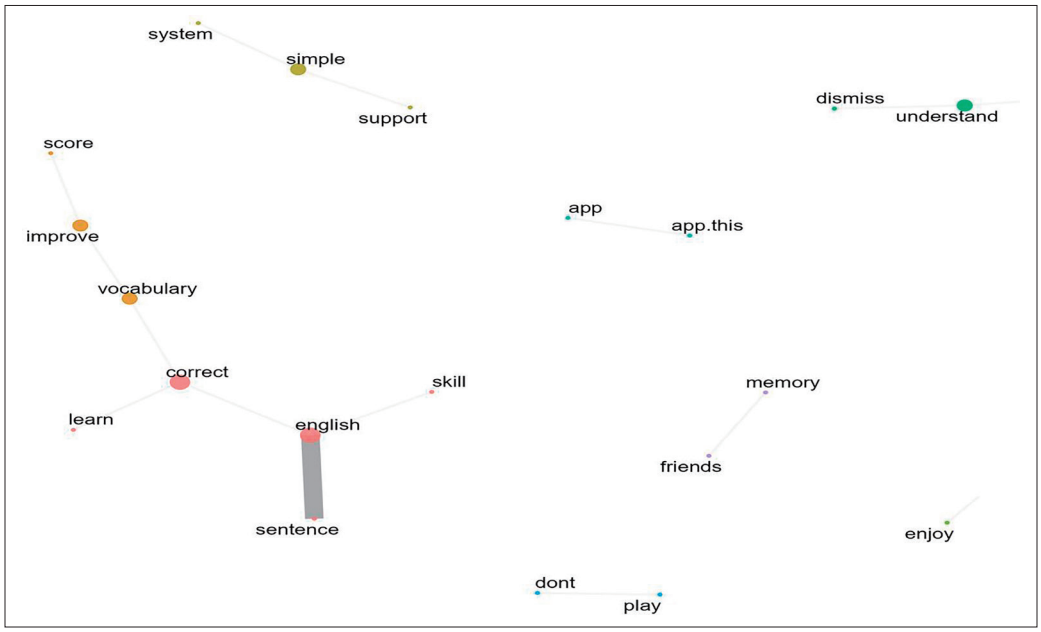
Table 4 Written Communication Questionnaire Results of all Apps Q3-5 Yes/No Answers

3. Did you enjoy using the app?			
Grammarly	Wordtune	ProwritingAid	QuillBot
Yes - 95/104	Yes - 98/117	Yes - 93/105	Yes - 90/105
91.35%	83.76%	88.57%	85.71%
4. Would you use this app again to improve your English writing skills?			
Grammarly	Wordtune	ProwritingAid	QuillBot
Yes - 87/104	Yes - 94/117	Yes - 77/105	Yes - 83/105
83.65%	80.34%	73.33%	79.05%
5. Did you find this app useful for developing and evaluating your writing skills today?			
Grammarly	Wordtune	ProwritingAid	QuillBot
Yes - 95/104	Yes - 89/117	Yes - 93/105	Yes - 85/105
91.35%	76.07%	88.57%	80.95%

Note. This table was created by the author using Word after manually counting questionnaire data.

Figure 3, generated in RStudio, shows a positive sentiment thematic network for Grammarly Q3. English and correct appeared most often with words like skill, learn, and sentence, while terms such as vocabulary and simple were next most frequent, linked to words like support and improve. Questionnaire responses support these findings, stating "It teaches my mistake word. I know the new word and Grammar" and "I was able to think again about the value my friends." These findings support research questions 2 and 3, in that students perceived AI as effective in improving English skills and supporting emotional needs.

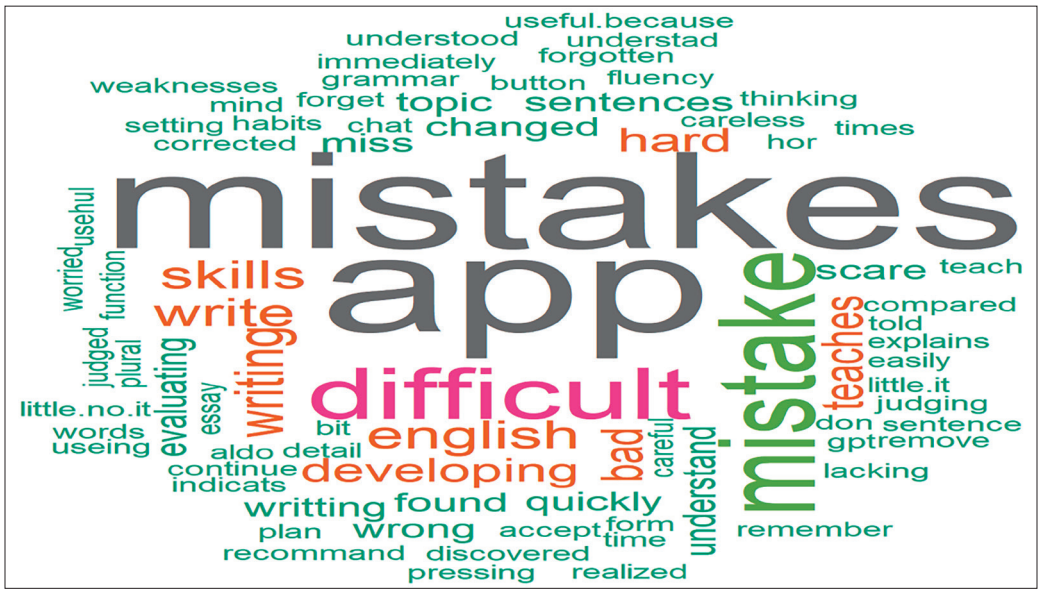
Figure 3 Positive Sentiment Thematic Network - Grammarly - Q3. Did you enjoy using the app?



Note. This figure was created by the author using RStudio (Posit Software, 2025).

Figure 4 displays a word cloud generated by RStudio, illustrating negative student sentiment towards Q5. Findings suggest that students did not find the apps useful, given that they made many mistakes, were difficult to use, and bad/hard for developing their skills. Supporting this, questionnaires said, "The setting changed immediately after pressing the accept button, making it difficult to tell what had changed" and "I couldn't understand why it was judged mistake."

Figure 4 Negative Sentiment Word Cloud - All Apps - Q5. Did you find this app useful for developing and evaluating your writing skills today?



Note. Images in this figure were created by the author using RStudio (Posit Software, 2025).

Interviews

From thirteen interviews, 220 codes were generated and organized into four categories/themes, as seen in Table 5. The data revealed that students enjoyed the AI tools primarily for their user-friendliness. ChatGPT and Grammarly were the most favoured, valued for their ability to improve communication skills and for emotional support. Aligned with the AI Directive's aim of preventing overreliance on AI, Bill emphasized the importance of self-regulated learning. He valued the personal attention AI provided, which addressed his writing weaknesses while offering tailored homework to meet his needs. Students reported that AI listened attentively and provided affirming feedback, thereby supporting learner autonomy and social-emotional needs.

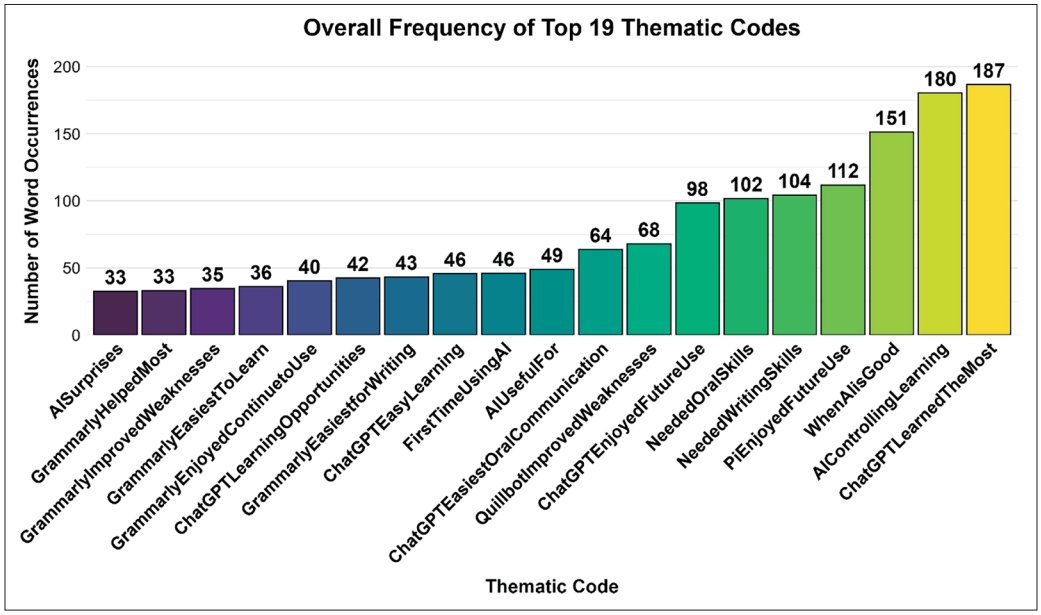
Table 5 Categories / Themes Created from Interview Coding

Categories / Themes					
Aliases Used for Participants		**Favorite AI Tools for Communicative Learning - Oral: 1st ChatGPT 2nd Pi Written: 1st Grammarly 2nd ProwritingAid**	**User-Friendly AI Tools are Perceived as More Effective Learning Tools!**	**Personalization Supports Emotional Psychological Needs, Promoting Long-Term Learning**	**Enjoyment Fun and Support Provided by AI Tools Motivated Learners to Improve their English Skills**
	Sophia	ChatGPT easiest to learn speaking. Grammarly was most enjoyable; ProwritingAid was the easiest to learn from.	ProwritingAid easiest to learn from, screen is very easy to understand writing skills evaluations.	Said AI writing apps should be tailored to users' characters and writing purposes, such as love letters, business, etc.	**Sue** liked Pi because it listened well, connecting its responses to what she said, allowing for smooth communication.
	Emma	Favoured ChatGPT - Grammarly helped most with writing and vocabulary.	Oral apps responded quickly and seemed to understand her thoughts, providing related questions that provoked fun and interest.	A perfect speaking app should utilize simple, user-directed vocabulary, grammar, and speed tailored to the user's level.	Enjoyed using Grammarly, as it quickly pointed out her grammar and spelling mistakes, motivating her to use it for future writing.
	Bill	ChatGPT helped him learn most. - Grammarly helped most with grammar.	Grammarly told his mistakes with difficult words and provided easier words that he happily remembers.	ChatGPT told him honestly he was not good at grammar and vocabulary and then provided personalized homework.	AI benefits students; it' s fun even when new, and they'll keep using it for English after the novelty fades.
	James	He enjoyed using Pi, but ChatGPT was first choice. ProwritingAid helped most with word sense.	ChatGPT helped with listening skills, because speed of talking is so clear and easy to listen to.	**Mike** said he asked Pi for English help and it clearly told him what and how to practise.	**Selena** loved using Pi—it complimented her, made her happy, and was easy to understand.

Note. This table was created by the author using Word, after manually coding and categorizing interview data.

In addition to 220 manual codes, 74 more were created with associated words, compiled into a CSV spreadsheet and analyzed in RStudio to produce the frequency bar chart in Figure 5. 'ChatGPTLearnedTheMost' was the most frequent code, and among the top 19 codes, ChatGPT and Grammarly appeared most often (five occurrences each), aligning with the highest yes responses in Tables 2 and 4. Each code was associated with various words spoken by the interviewees, with "ChatGPTLearnedTheMost" being associated with "not, good, vocabulary, grammar, said, you, are, not, good, and at." In comparison, "WhenAIisGood" was associated with "easy, fun, most important, pronunciation, listen, writing, word, sense, correct, and answer." Interviewees illustrate some of these words, for Emma said, "These apps understand my thoughts" and "I want to challenge, for EIKEN. So, I want to use this GPT for interview." Mike said, "I often forgot to write "A" or "The". So, I could know this is my weakness" and "I talked to, I asked ChatGPT, when we made a big mistake, what do you, what should we think of? … Making mistakes is process of success."

Figure 5 Overall Frequency of the Top 19 Thematic Codes Created from 74 total codes.



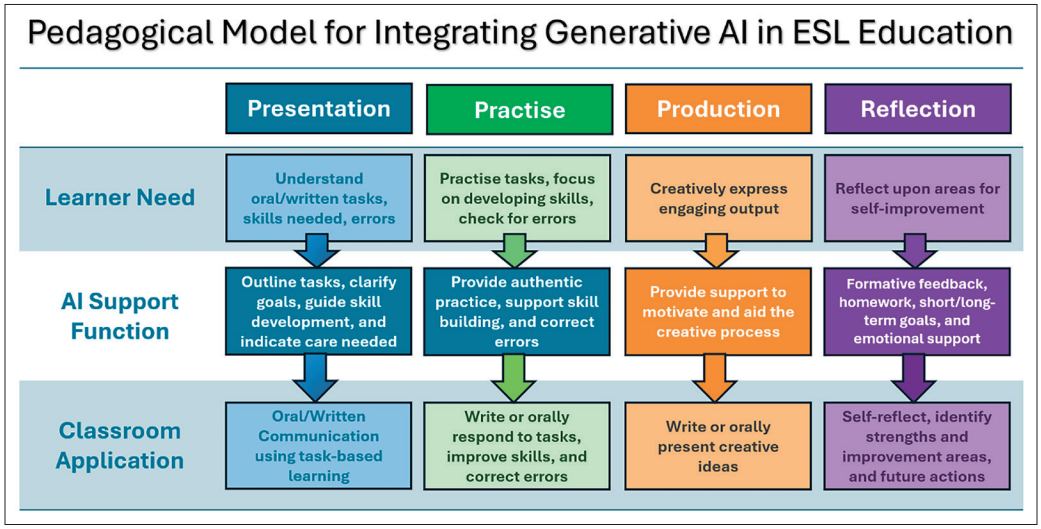
Note. This figure was created by the author using Thematic Data Visualization in RStudio (Posit Software, 2025).

Conclusion

Some might argue, and rightfully so, that the findings of this study may be dated, given that it was conducted in 2024 and that AI models have changed drastically since then. Although these factors are indisputable, the pedagogical model employed in this study, particularly the AI Directive-tailored prompts and in-class processes, ensures that the findings remain robust and applicable across current and future AI models.

In this study, students preferred ChatGPT and Grammarly because they were perceived as user-friendly and effective in providing immediate, positive, and tailored feedback that supported the development of oral and written communication skills, fostered autonomy and kindness, and promoted long-term intrinsic motivation. These findings support the AI Directive and the prompts used in this study, as they focus on providing AI-based learning that supports autonomous student learning, emotional support, and social and interactive growth. Even as AI models evolve, the pedagogical approach used in this study continues to support meaningful student interactions that deliver personalized feedback and targeted homework to address English-language weaknesses, while fostering intrinsic motivation. Table 6 illustrates the approach used in this study as a pedagogical model for integrating generative AI into ESL education, which leverages the support functions inherent in generative AI tools to meet learners' needs.

Table 6 Pedagogical Model for Integrating Generative AI in ESL Education.



Note. This table was created by the author using Microsoft PowerPoint.

Study limitations include potential student bias toward popular apps, technical issues such as Wi-Fi and device/app malfunctions, and the fixed weekly order of apps, which may lead to order effects as students improve their AI use, become more familiar with it, and adjust their comparison criteria. Additionally, because responses were collected anonymously and could not be linked across tools at the participant level, within-subject statistical comparisons across tools were not possible. Therefore, tool comparisons are reported descriptively as response patterns rather than as inferential tests of differences. Finally, future research should examine customizable AI environments (e.g., MyGPTs/Gems) to determine whether tailored prompts and tool configurations change learners' experiences and perceptions.

References

Abe, E. (2013). Communicative language teaching in Japan: Current practices and future prospects: Investigating students' experiences of current communicative approaches to English language teaching in schools in Japan. *English Today*, *29*(2), 46-53. https://doi.org/10.1017/S0266078413000163

Ahmed, I., Ghafoor, A., Liliuara, D., & Akyuningrum, V. Q. (2025). The impact of AI learning tools on ESL learners' motivation and success: A systematic literature review. *English Learning Innovation (Englie)*, *6*(1), Article 1. https://doi.org/10.22219/englie.v6i1.38961

Aizawa, I., & Rose, H. (2020). High school to university transitional challenges in English medium instruction in Japan. *System*, *95*, 102390. https://doi.org/10.1016/j.system.2020.102390

Alharbi, W. (2023). AI in the foreign language classroom: A pedagogical overview of automated

writing assistance tools. *Education Research International*, *2023*, 1-15. https://doi.org/10.1155/2023/4253331

Almelhes, S. A. (2023). A review of artificial intelligence adoption in second-language learning. *Theory and Practice in Language Studies*, *13*(5), 1259-1269. https://doi.org/10.17507/tpls.1305.21

Busso, A., & Sanchez, B. (2024). Advancing communicative competence in the digital age: A case for AI tools in Japanese EFL education. *Technology in Language Teaching & Learning*, *6*(3), 1211. https://doi.org/10.29140/tltl.v6n3.1211

Capinding, A. T., & Dumayas, F. T. (2024). Transformative pedagogy in the digital age: Unraveling the impact of artificial intelligence on higher education students. *Problems of Education in the 21st Century*, *82*(5), 630-657. https://doi.org/10.33225/pec/24.82.630

CASEL (2023). *What is the Casel framework*? https://casel.org/fundamentals-of-sel/what-is-the-casel-framework/

Chauncey, S. A., & McKenna, H. P. (2023). A framework and exemplars for ethical and responsible use of AI chatbot technology to support teaching and learning. *Computers and Education: Artificial Intelligence*, *5*, 100182. https://doi.org/10.1016/j.caeai.2023.100182

Chen, D., Liu, W., & Liu, X. (2024). What drives college students to use AI for L2 learning? Modeling the roles of self-efficacy, anxiety, and attitude based on an extended technology acceptance model. *Acta Psychologica*, *249*, 104442. https://doi.org/10.1016/j.actpsy.2024.104442

Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence*, *6*, 100230. https://doi.org/10.1016/j.caeai.2024.100230

EF Epi (2024). *EF English Proficiency Index*. https://www.ef.com/wwen/epi/

Guo, Y., & Wang, Y. (2024). Exploring the effects of artificial intelligence application on EFL students' academic engagement and emotional experiences: A mixed-methods study. *European Journal of Education*, *60*(1), e12812. https://doi.org/10.1111/ejed.12812

Iwamoto, N. (2016). An investigation of Japanese university students' English-speaking skills. *Journal of the Association of English Language Teachers of Japan* 26: 1-16.

Jalleh, C. M., Mahfoodh, O. H., & Singh, M. K. (2021). Oral communication apprehension among Japanese EFL international students in a language immersion program in Malaysia. *International Journal of Instruction*, *14*(2), 155-178. https://doi.org/10.29333/iji.2021.14210a

Kohnke, L., & Moorhouse, B. L. (2025). Enhancing the emotional aspects of language education through generative artificial intelligence (GenAI): A qualitative investigation. *Computers in Human Behavior*, *167*, 108600. https://doi.org/10.1016/j.chb.2025.108600

Liu, M., & Reinders, H. (2025). Do AI chatbots impact motivation? Insights from a preliminary longitudinal study. *System*, *128*, 103544. https://doi.org/10.1016/j.system.2024.103544

Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S. (2024). Exploring the application of

ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments*, *11*(1), 1-24. https://doi.org/10.1186/s40561-024-00342-5

Long, R., & Watanabe, H. (2021). The mirage of progress? A longitudinal study of Japanese students' L2 oral grammar. *The Language Teacher*, *45*(2), 9. https://doi.org/10.37546/JALTTLT45.2-2

Mizumoto, A., Shintani, N., Sasaki, M., & Teng, M. F. (2024). Testing the viability of ChatGPT as a companion in L2 writing accuracy assessment. *Research Methods in Applied Linguistics*, *3*(2), 100116. https://doi.org/10.1016/j.rmal.2024.100116

Morita, L. (2017). Why Japan needs English. *Cogent Social Sciences*, *3*(1), 1399783. https://doi.org/10.1080/23311886.2017.1399783

Nuske, K. (2019). Vehicle of erotic liberation or instrument of career survival? Japan's ideologies of English as reflected in conversation school advertisements. *Discourse, Context & Media*, *31*, 100319. https://doi.org/10.1016/j.dcm.2019.100319

Pfau, A., Polio, C., & Xu, Y. (2023). Exploring the potential of ChatGPT in assessing L2 writing accuracy for research purposes. *Research Methods in Applied Linguistics*, *2*(3), 100083. https://doi.org/10.1016/j.rmal.2023.100083

Posit Software, PBC. (2025). *RStudio* (Version 2025.09.1, Build 401) [Computer software]. https://posit.co/products/open-source/rstudio/

Sakamoto, M., & Furukawa, G. (2022). Native speakers aren't perfect': Japanese English learners' identity transformation as English users. *System*, *110*, 102921. https://doi.org/10.1016/j.system.2022.102921

Samuell, C. (2021). Learner attitudes towards English: considering the Japanese context. *Hannan Ronshū Shakai Kagakuhen*, *57*(1), 21-33.

Sharadgah, T. A., & Sa'di, R. A. (2022). A systematic review of research on the use of artificial intelligence in English language teaching and learning (2015-2021): What are the current effects? *Journal of Information Technology Education: Research*, *21*, 337-377. https://doi.org/10.28945/4999

Smith, C. (2025). Challenges in English language education in Japan: Policy vs. practice. *European Journal of Education and Pedagogy*, *6*(2), 53-63. https://doi.org/10.24018/ejedu.2025.6.2.912

Son, J.-B., Ružić, N. K., & Philpott, A. (2023). Artificial intelligence technologies and applications for language learning and teaching. *Journal of China Computer-Assisted Language Learning*, *5*(1), 94-112. https://doi.org/10.1515/jccall-2023-0015

Takiguchi, M., & Machida, J. (2020). The current situation and issues with English language activities in primary schools in Japan after six years of implementation. *Vietnam Journal of Education*, 14-24. https://doi.org/10.52296/vje.2020.3

Terasawa, T. (2022). What made primary English education in Japan different from the global

trend? A policy process analysis. *Linguistics and Education*, *71*, 101084. https://doi.org/10.1016/j.linged.2022.101084

UNESCO (2022). *Recommendation on the ethics of artificial intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000381137

UNESCO (2024). *Mainstreaming social and emotional learning in education systems: Policy guide*. https://doi.org/10.54675/ORWD6913

Wang, Y. (2025). A study on the efficacy of ChatGPT-4 in enhancing students' English communication skills. *Sage Open*, *15*(1), 1-17. https://doi.org/10.1177/21582440241310644

Wicaksono, M., & Saraswati, N. (2024). Social emotional learning in ELT: Opportunities and challenges. *Global Expert: Jurnal Bahasa Dan Sastra*, *12*(1), 25-33. https://doi.org/10.36982/jge.v12i1.4427

Wu, X.Y. (2024). Artificial intelligence in L2 learning: A meta-analysis of contextual, instructional, and social-emotional moderators. *System*, *126*, 103498. https://doi.org/10.1016/j.system.2024.103498

Wu, X., & Li, R. (2024). Unraveling effects of AI chatbots on EFL learners' language skill development: A meta-analysis. *The Asia-Pacific Education Researcher*. Advance online publication. https://doi.org/10.1007/s40299-024-00853-2

Yamaoka, K. (2024). ChatGPT's motivational effects on Japanese university EFL learners: A qualitative analysis. *International Journal of TESOL Studies*, *6*(3). 24-35. https://doi.org/10.58304/ijts.20240303

Yang, H., & Kyun, S. (2022). The current research trend of artificial intelligence in language learning: A systematic empirical literature review from an activity theory perspective. *Australasian Journal of Educational Technology*, 180-210. https://doi.org/10.14742/ajet.7492

Zong, Y., & Yang, L. (2025). How AI-enhanced social-emotional learning framework transforms EFL students' engagement and emotional well-being. *European Journal of Education*, *60*(1), e12925. https://doi.org/10.1111/ejed.12925

Appendix A
Prompts Used for Oral and Written Communication Apps

"Please engage in an oral conversation at about an EIKEN 2 level with me, talking about _____, for a duration of 10 minutes. Give me about five seconds after you ask your question for me to think. I'm looking forward to a meaningful conversation with you."

"At the end of our 10 minutes, please stop and assess my oral communication based on the following criteria, each with a possible rating of 100%: Pronunciation, Fluency, Vocabulary, Grammar, and Coherence and Cohesion. Afterwards, provide me with constructive feedback and actionable recommendations for improvement."

"In addition to evaluating my communication skills, please consider providing supportive and encouraging feedback to address my social and emotional needs. Your guidance in fostering self-esteem, confidence, and motivation for further growth is greatly appreciated."

"As part of our conversation, I'd like to receive a homework assignment that aligns with the insights we've discussed. Please provide me with a task that will help me proactively apply the feedback and recommendations you've offered."

"Thank you for our conversation. Before we conclude, I'd like to receive motivation and encouragement to take proactive steps based on the feedback and homework assignments. Inspire me to continue developing my communication skills and exploring new interests."

"Please evaluate a piece of my writing by giving me scores out of 100% for each of the following criteria: vocabulary, grammar, content, organization, and coherence and cohesion.
Again, please be sure to give each criterion a score out of 100%.

"After giving me my writing evaluation, please provide me with constructive feedback and actionable recommendations for improvement."

"Besides evaluating my writing skills, please consider providing supportive and encouraging feedback to address my social and emotional needs. Your guidance in fostering self-esteem, confidence, and motivation for further growth is greatly appreciated."

"As part of our conversation, I'd like to receive a homework assignment that aligns with the insights you have provided. Please provide me with a task that will help me proactively apply the feedback and recommendations you've offered."

"Here is my written work to evaluate:"

Appendix B

Content Used for Google Forms Questionnaire for Oral Communication Apps

Please Check Which App You Used Today

☐ Gemini ☐ Bing ☐ ChatGPT ☐ Pi

1. Please give up to 100% for your scores starting with

a) Pronunciation:

b) Fluency:

c) Vocabulary:

c) Grammar:

D) Coherence and Cohesion

2. How easy was it to understand how to use the app? Please rate from 1 (Very Easy) to 7 (Very Difficult)

3. Did you enjoy using the app? Please explain.

4. Would you use this app again to improve your English or simply to have fun? Please explain.

5. Did you find this app useful for discussing today's topic or question? Please explain...

Appendix C

Content Used for Google Forms Questionnaire for Written Communication Apps

Please Check Which App You Used Today

☐ ProwritingAid ☐ Wordtune ☐ Grammarly ☐ QuillBot

1. Please give up to 100% for your scores starting with

a) Vocabulary:

b) Grammar:

c) Content:

c) Organization:

D) Coherence and Cohesion

2. How easy was it to understand how to use the app? Please rate from 1 (Very Easy) to 7 (Very Difficult)

3. Did you enjoy using the app? Please explain.

4. Would you use this app again to improve your English writing skills? Please explain.

5. Did you find this app useful for developing and evaluating your writing skills today? Please explain...

Appendix D
Interview Guide

Part A - Oral Communication Apps

1. What do you think are the most important skills to be a good English speaker?
2. Did any of the four AI oral communication apps help you with all or most of the skills you just mentioned? Please explain.
3. Which of the four AI oral communication apps did you think were easiest to use and to learn from? Please explain.
4. Which AI oral communication app did you enjoy using the most and perhaps will continue to use in the future? Please explain.
5. Which AI oral communication app did you think you learned the most from, and did it help guide you on how to improve your weaknesses? Please explain.

Part B - Written Communication Apps

6. What do you think are the most important skills to be a good English writer?
7. Did any of the four AI writing communication apps help you with all or most of the skills you just mentioned? Please explain.
8. Which of the four AI writing communication apps did you think were easiest to use and to learn from? Please explain.
9. Which AI writing communication app did you enjoy using the most and perhaps will continue to use in the future? Please explain.
10. Which AI writing communication app did you think you learned the most from, and did it help guide you on how to improve your weaknesses? Please explain.

If time permits:

11. Did you find you could control your own learning and that the AI tools were simple, helpful tools, or did you feel the learning process was fully controlled by the AI tools? Please explain.
12. How do you think we can know if an AI app is good for educating our students?