
RESEARCH ARTICLE

Exploring GenAI Chatbot Use for L2 Learning: A Mixed-Methods Study of Foundation-Year Students in Oman

Moosa Ahmed Bait Ali Sulaiman¹, Mohamed AboulEla Abdelmageed², Nagwa Mohammed Khallaf³

¹Director of Centre for Preparatory Studies and Associate Professor of English, Dhofar University, Salalah, Oman. ORCID: <https://orcid.org/0000-0002-3775-6165>

²English Language Lecturer at Centre for Preparatory Studies, Dhofar University, Salalah, Oman. ORCID: <https://orcid.org/0000-0003-3165-0711>

³Director of Centre for Sustainability and Assistant Professor of English, Dhofar University, Salalah, Oman. ORCID: <https://orcid.org/0009-0001-0034-7544>

Corresponding Author: Nagwa Mohammed Khallaf E-mail: (nkhallaf@du.edu.om) & (nagwakhallaf2010@gmail.com)

ORCID: <https://orcid.org/0009-0001-0034-7544>

ABSTRACT

Generative artificial intelligence (GenAI) chatbots have increasingly become part of many university students' English as a second language (L2) learning. However, little is known about how students in general foundation programmes use them during the move from Arabic-medium schooling to English-medium higher education. This sequential explanatory mixed-methods study examined GenAI chatbot use among foundation-year students at a private university in Oman. A questionnaire (N = 262) covered usage patterns, perceived usefulness for L2 skills, AI output verification practices, challenges and support needs, and was followed by two focus groups (n = 15). Results showed that ChatGPT was the dominant tool (91.2%) that students accessed mainly by smartphone (80.5%). The most common frequency was "sometimes" (47.3%), while 42.4% of respondents reported using chatbots "always" or "often". Most students (57.6%) switched between English and Arabic while interacting with GenAI chatbots. Students rated chatbots as clearly useful for writing, speaking and reading, but not for listening. They also valued grammar and vocabulary corrections by chatbots more than their feedback on writing or speaking assignments. More than two-thirds of respondents preferred teacher feedback on their writing (72.9%). The most frequently reported challenges were incorrect information (43.5%) and unclear responses (35.1%), while the most common verification strategy of generated content was asking a teacher (46.2%). Focus-group participants described purpose-driven but largely unplanned prompting, translation-mediated comprehension, concern about overreliance, inconsistent teacher stances, and a need for classroom-embedded guidance. Chatbot use for L2 learning among foundation-year students can be interpreted as a tool-mediated activity largely outside formal L2 instructional design, with teachers as epistemic anchors. The study recommends early, skill-specific, teacher-mediated AI literacy instruction in foundation programmes.

KEYWORDS

Generative Artificial Intelligence (GenAI), AI Chatbots, English as a Second Language Learning, Foundation Programme, AI Literacy, Oman.

ARTICLE INFORMATION

ACCEPTED: 02 September 2026

PUBLISHED: 16 October 2026

DOI: 10.32996/jeltal.2026.8.11.1

1- Introduction

Generative artificial intelligence (GenAI) chatbots are conversational interfaces to large language models, such as ChatGPT, Gemini and Copilot. They can generate text and, increasingly, speech in response to natural-language prompts. "GenAI chatbots" (or

simply "chatbots") refers to this class of tools, while "ChatGPT" refers only to that product. For second language (L2) learners, these tools offer on-demand explanation, text generation, correction and conversational practice (Kohnke et al., 2023). Meta-analyses have reported medium positive effects of chatbots, including GenAI chatbots, on L2 learning (Li et al., 2025; Lyu et al., 2025; Wang et al., 2025). Perception studies describe ChatGPT as a "companion" for writing feedback (Teng, 2024). However, these findings show more about what chatbots can do in research settings than about how L2 learners actually use them.

Two constraints bound how far these findings transfer to L2 learning. First, effect studies typically examine interventions designed by researchers, whereas most students' chatbot use is self-initiated, occasional and largely invisible to teachers. Much less is known about what learners actually do with chatbots and how this relates to L2 instruction (Yang & Kyun, 2022). Second, the evidence base is geographically uneven. Gulf research is growing, but it has mostly surveyed acceptance and perceptions among general undergraduate or credit-course populations (Abdaljaleel et al., 2024; Alshamy et al., 2025; Ali et al., 2025). Both limitations matter most where chatbot habits are still forming in foundation programmes.

General Foundation Programmes (GFPs) in Oman, which prepare students for effective engagement in higher education (Oman Academic Accreditation Authority [OAAA], 2008) and have been externally audited since 2017 (OAAA, 2017), are a distinctive and consequential setting. Mandatory in all Omani higher education institutions since 2009/2010, GFPs typically last one year and cover English, mathematics, computing and general study skills (Elamin, 2025). Students enter after Arabic-medium secondary schooling with the aim of reaching, within that year, the English proficiency required for English-medium degree study, an exit standard of IELTS band 5.0 with no skill below 4.5 (Al-Issa & Al-Bulushi, 2012), roughly upper B1 (Elamin, 2025). Institutional practices for these entry and exit standards remain largely unsystematic (Al-Issa & Al-Bulushi, 2012; Al-Rusheidi, 2021). Their chatbot habits therefore form at the same time as their academic literacy practices (see Aypay & Bedeker, 2025). In this setting, students' language choices, devices, generated content verification strategies and reliance matter as much as their satisfaction ratings.

Students in Omani higher education use GenAI chatbots for academic support, translation, grammar checking and idea generation (Al-Raimi et al., 2024; Alshamy et al., 2025; Ali et al., 2025), while institutional guidance has not kept pace (Benayoune et al., 2026). The closest Omani studies examined foundation students' views of ChatGPT (Dehdary & Meschi, 2024) and their perspectives on AI tools for writing (Al-Raimi et al., 2024), while international reviews show that research has concentrated on writing and on outcomes such as writing quality (Crosthwaite & Sun, 2026; Lo et al., 2024). To the best of the researchers' knowledge, no study has provided a systematic account of how often GFP students use chatbots, for what purposes, in which language, with what checks, or in what relation to classroom teaching. Therefore, this study addresses that gap through four research questions.

1-1- Research Questions

The key research questions include:

- How often and for what purposes do foundation-year students use GenAI chatbots?
- How useful do students find GenAI chatbots across L2 skills and for feedback?
- How do students position their chatbot use in relation to classroom instruction and their teachers?
- What challenges, verification practices and support needs do students report?

In so doing, the study documents chatbot use patterns among GFP students, interprets them as mediated activity not yet part of formal L2 instructional design, and identifies concrete targets for AI literacy instruction in foundation programmes.

2- Literature Review

2-1-1. What chatbots can and cannot do for L2 learning

Before the advent of ChatGPT, a systematic review of 25 studies identified technological, pedagogical and social affordances of chatbots for L2 learning. The same review also identified technological limitations, novelty effects and cognitive load as recurring problems (Huang et al., 2022). Chatbots based on large language models have extended these affordances to open-ended text generation, explanation and adaptive dialogue (Kohnke et al., 2023). Whether these affordances translate into learning gains, however, is less clear.

Broadly, aggregate effects are positive but vary considerably. For example, Wang et al. (2025) reported a moderate overall effect ($g = 0.484$) across 28 studies, moderated by educational level, language proficiency, interface design and interaction capability. Two further meta-analyses have confirmed medium effects for L2 learning (Li et al., 2025; Lyu et al., 2025): $g = 0.61$ across 31 comparative studies and 0.58 across 41 GenAI-chatbot studies, respectively. However, the robustness of these syntheses is contested. Auditing the experimental studies behind them, Weidlich et al. (2025) found that few clearly described the treatment, specified what control groups did, or measured durable learning. As such, the reported gains cannot be confidently attributed to the chatbots themselves. Additionally, these syntheses mainly aggregate short, researcher-designed interventions and do not reveal how learners use chatbots on their own initiative.

2-1-2. Skill-specific evidence: the dominance of writing

Most empirical L2 research on GenAI concerns writing (Lo et al., 2024). Barrot (2023) highlighted the potential of ChatGPT for idea generation, feedback and attention to language form despite risks for L2 writing pedagogy and academic integrity. Furthermore, Escalante et al. (2023) found that AI-generated feedback produced writing gains comparable to those from human tutor feedback. A scoping review of 51 SSCI-indexed studies of GenAI written feedback found that most measure gains in writing quality through revision analyses, pre- and post-tests and surveys (Crosthwaite & Sun, 2026). This leaves learners' self-directed chatbot use largely unexamined. Benefits may also be short-lived: among Iranian B1 English majors, ChatGPT-supported essay writing raised task motivation, but its advantage over peer collaboration faded by a one-month follow-up (Zare et al., 2025). These mixed findings echo the three contradictions that Warschauer et al. (2023) theorised for L2 writers: imitation (AI-generated text aids learning but risks plagiarism), rich-get-richer (benefits depend on unequal access and skill) and with-or-without (using AI well requires the very writing skills that AI use may short-circuit). These contradictions frame the prompting practices examined in this study.

Research on L2 speaking and listening is smaller but growing. In a quasi-experimental study, AI-mediated interaction improved English as a foreign language (EFL) learners' speaking performance and willingness to communicate (Fathi et al., 2024). This scarcity may partly reflect the text-based mode in which GenAI chatbots have mostly been used. However, current voice-enabled chatbots allow spoken interaction, and chatbot modality and interaction capability moderate learning effects (Lyu et al., 2025; Wang et al., 2025). Listening has rarely been examined directly, so skill coverage remains uneven.

2-1-3. Learner perceptions and use in comparable contexts

Perception studies have shown that L2 learners value chatbots for grammar, vocabulary, explanation and idea generation in spite of their concerns about accuracy and overreliance (Chan & Hu, 2023; Xiao & Zhi, 2023). Such views have also surfaced outside formal education. In 1,088 comments on popular YouTube videos about ChatGPT for language learning, learners cited practice opportunities, individualised feedback and real-time linguistic help as benefits (Chen et al., 2026). In the Arab region, a five-country survey of 2,240 university students found that 46.8% had heard of ChatGPT and 52.6% had used it. Ease of use, usefulness, social influence, low perceived risk and low anxiety predicted positive attitudes (Abdaljaleel et al., 2024). Whether such patterns hold for Omani foundation-year students is less clear.

Previous Omani have rarely isolated foundation-year students. At Sultan Qaboos University, for instance, students reported frequent GenAI use for academic support, including completing assignments (Alshamy et al., 2025). Among 61 preparatory-studies learners at the University of Technology and Applied Sciences in Salalah, translation was the most common use of AI tools, followed by spelling and grammar checks and idea generation (Al-Raimi et al., 2024). Among 198 Omani learners at two institutions, ChatGPT served for explanations, exercise answers, grammar and ideas despite concerns about overreliance, accuracy and integrity (Ali et al., 2025). Dehdary and Meschi (2024) reported broadly positive perceptions among Omani foundation and credit-course students, together with concerns about cheating and dependency. In a classroom study at an Omani public university, students who were first trained in prompting and ethical use reported positive effects on writing, speaking and engagement (Jahan & Akbar, 2026). Beyond the Gulf, foundation-year students at an English-medium university in Kazakhstan used ChatGPT as a "literacy advisor" while questioning its limits (Aypay & Bedeker, 2025). These studies describe what students use chatbots for but rarely how they check outputs.

2-1-4. Research gap

Three contentions run through this literature. The first concerns benefit versus cognitive offloading. A systematic review associates overreliance on AI dialogue systems with weaker decision-making, critical thinking and analytical reasoning (Zhai et al., 2024). Experimental evidence points in the same direction. In a field experiment with nearly 1,000 secondary-school students, unrestricted access to a GPT-4 assistant improved performance on practice problems but lowered subsequent unassisted exam performance, whereas a version designed with tutoring safeguards largely avoided this harm (Bastani et al., 2025). In a randomised study of university students' essay writing, ChatGPT support improved essay scores but produced no advantage in knowledge gain or transfer and was associated with less metacognitive activity (Fan et al., 2025). Though neither involved L2 learners, both studies suggest that how a chatbot is used and with what safeguards matter more than access per se. The second tension concerns replacement versus complementarity. Most empirical work supports complementarity, with teachers contributing pedagogical judgement, relational support and curricular fit that chatbots do not provide (Jeon & Lee, 2023). The third tension is between acceptance and practice. Much Gulf research measures attitudes and intention to use but does not extensively tackle interaction language choice, GenAI content verification, or relation to L2 instruction (Alshamy et al., 2025; Ali et al., 2025).

Methodologically, self-report surveys of general university samples predominate (e.g., Abdaljaleel et al., 2024; Ali et al., 2025; Chan & Hu, 2023). Besides, foundation-year students in Oman appear only within mixed samples that have rarely examined L1 use in interaction with chatbots, access barriers, and output verification (Al-Raimi et al., 2024; Dehdary & Meschi, 2024). The present study addresses these gaps for a foundation-year population.

3- Conceptual Framework

This study adopts a mediational framing rather than a technology acceptance model because it measures practices and perceptions, not acceptance constructs. Following activity-theoretical work on AI in language learning (Yang & Kyun, 2022), chatbots are treated as mediating tools within learners' activity systems. Whether they support learning depends on learners' goals, the tasks they are set, the rules that govern use (for example, teacher permissions) and the division of labour among learner, tool and teacher. Godwin-Jones (2024) extends this view by describing agency in GenAI-supported language learning as distributed between learner and machine. Who does the cognitive and linguistic work therefore becomes an empirical question.

Furthermore, AI literacy is the competence that makes such mediation productive. Building on Long and Magerko's (2020) foundational definition, recent syntheses describe it as the ability to recognise, understand, use and critically evaluate AI, and to navigate its use ethically (Almatrafi et al., 2024; Miao & Shiohira, 2024; Yang et al., 2025). In L2 writing specifically, it has been operationalised as understanding and applying AI, awareness of its ethical risks, and self-efficacy in AI-assisted writing (Huang et al., 2025). Critical digital literacies approaches extend the construct by adding awareness of how GenAI outputs shape voice, authority and ideology (Darvin, 2025). Such competence cannot be assumed, however, because AI literacy varies widely among L2 learners. For example, a survey of 1,001 Chinese EFL writers identified four profiles, ranging from "skeptical and resistant" to "comprehensive high-literacy", which differed in writing self-regulation (Pan & Zhang, 2026). Together, mediation and AI literacy yield three questions that guide the interpretation of the findings: which resources (language, devices, teachers) mediate students' chatbot use; who does the cognitive and linguistic work in learners' chatbot interactions; and which AI literacies students display or lack.

4- Materials and Methods

4-1- Design and setting

The study used a sequential explanatory mixed-methods design, with a cross-sectional questionnaire followed by focus groups. It was conducted in the Foundation Program of a private Omani university during the spring term of the 2024–2025 academic year. The programme prepares students, the majority of whom come from Arabic-medium secondary schools, for English-medium undergraduate study. It has three levels (Levels 1–3), to which students are assigned by the online Cambridge English placement test.

4-2- Participants

Approximately 1,000 students were enrolled in the programme at the time of data collection. Using convenience sampling across all three levels, 262 students completed the questionnaire, a response rate of 26.2%. Because sampling was non-probabilistic, the results describe this sample rather than estimate population values.

As shown in Table 1, fifteen students (seven male, eight female) took part in two focus groups: six from Levels 1 and 2 (three male, three female) and nine from Level 3 (four male, five female). Teachers recommended participants who were articulate, willing to share their experiences and representative of diverse perspectives. Teachers confirmed that focus group participants completed the anonymous questionnaire.

Table (1)
Participant Characteristics (N = 262)

Characteristic	Category	n	%
Gender	Female	132	50.4
	Male	130	49.6
Age	18–20	230	87.8
	21–23	17	6.5
	24 or above	15	5.7
Foundation level	Level 1	47	17.9
	Level 2	58	22.1
	Level 3	157	59.9
Years of English study	Less than 5	67	25.6
	5–8	25	9.5
	9–12	43	16.4
	More than 12	126	48.1
	Other	1	0.4

4-3- Instrument

The researchers-developed questionnaire (following DeVellis & Thorpe, 2022) had seven sections: (1) demographics; (2) background, covering how students first learned about chatbots, the chatbot used most, the access device and comfort level (5-point scale); (3) usage, covering frequency, language of interaction, change in use, session duration and purposes; (4) perceived effectiveness for reading, writing, listening and speaking (5-point scale, 1 = not effective to 5 = highly effective) and preferred feedback source; (5) verification methods and the perceived effectiveness of chatbots in correcting grammar and vocabulary and in giving feedback on writing and speaking; (6) challenges and support needs; and (7) four 5-point agreement items on perceived improvement, confidence, independence and willingness to recommend chatbots.

After piloting with 15 foundation students, who were excluded from the main sample, all items were presented in both English and Arabic. Three experts in applied linguistics, educational technology and research methods reviewed content validity, and items endorsed by at least two of the three were retained. The original analysis reported an overall Cronbach's α of .80 across all items. Because the questionnaire combines different item types, this value is reported for completeness rather than as evidence of a single construct. The skill-effectiveness items (Section 4) and the agreement items (Section 7) are analysed item by item rather than summed into scale scores, so subscale reliability is not required for the analyses reported here.

4-4- Procedures

The questionnaire was built in Microsoft Forms. Instructors shared the link with their classes, and students completed the questionnaire on their mobile phones. Participation was voluntary, and students gave electronic informed consent before beginning. After preliminary analysis of the questionnaire data, two semi-structured focus groups (Krueger & Casey, 2015) of about one hour each were held in a private room on campus with two facilitators. Participants could contribute in English or Arabic. The discussion guide was developed from the preliminary questionnaire results, and covered frequency and purposes of use, skill-specific benefits, strategies, classroom integration, challenges and recommendations. For example, students were asked why chatbots seemed less useful for listening and why they preferred teacher feedback. Sessions were audio-recorded and transcribed verbatim. Contributions in Arabic were translated into English by a member of the research team and checked by a second researcher.

5- Analysis

Questionnaire data were analysed in SPSS (version 30.0). Frequencies and percentages are reported for categorical and select-all items. For the four skill-effectiveness items and the four overall agreement items (5-point scales), means, standard deviations, 95% confidence intervals and the percentage of ratings of 4 or 5 are reported. One-sample t-tests against the scale midpoint of 3, checked with Wilcoxon signed-rank tests, show whether each rating departed from a neutral judgement. Agreement-item statistics were computed from their response distributions. Pairwise differences between skill ratings were tested conservatively: the paired t statistic was computed under the least favourable pairing of responses compatible with each pair's observed distributions, that is, the pairing with the lowest possible inter-item correlation. A statistically significant difference under this pairing is significant under any pairing; Bonferroni correction was applied across six comparisons. The same least-favourable-pairing logic was applied, using McNemar's test, to compare the proportions rating chatbot correction and chatbot feedback as effective. Preference for teacher versus chatbot feedback was examined with an exact binomial test. Effect sizes are reported as Cohen's d and interpreted following Cohen (1988). Section totals were not analysed, because they combine nominal, ordinal and select-all items.

Focus-group transcripts were analysed through inductive thematic analysis following Braun and Clarke's (2006) six phases. Two researchers independently coded 25% of the transcripts, compared their codes and resolved discrepancies through discussion until consensus was reached; the remaining transcripts were coded using the agreed codebook. The two strands were integrated at the interpretation stage by mapping qualitative themes onto quantitative patterns for each research question (see Table 6).

6- Results

6-1- Usage patterns (RQ1)

More than half of the respondents first learned about chatbots through social media (53.4%), followed by teachers (25.2%), classmates (10.7%), self-discovery (4.6%) and other sources (6.1%). ChatGPT was the most-used tool (91.2%), followed by Gemini (5.0%), Copilot (1.5%) and others (2.3%). Access was mainly by smartphone (80.5%), with tablets (11.5%) and laptops (8.0%) less common. Just over half of the students felt comfortable or very comfortable using chatbots for learning English (55.7%), 24.0% were neutral and 20.2% felt uncomfortable. Overall comfort was above neutral ($M = 3.63$, $SD = 1.33$; $t(261) = 7.70$, $p < .001$, $d = 0.48$). Table 2 illustrates usage patterns.

Table (2)
Usage Patterns (N = 262)

Item	Response	n	%
Frequency of use	Always	56	21.4
	Often	55	21.0
	Sometimes	124	47.3
	Rarely	19	7.3
	Never	8	3.1
Language of interaction	English only	60	22.9
	Arabic only	51	19.5
	Switch between English and Arabic	151	57.6
Change in use	Increased a lot	55	21.0
	Increased a little	71	27.1
	Stayed the same	102	38.9
	Decreased a little	26	9.9
	Decreased a lot	8	3.1
Typical session length	Less than 15 minutes	134	51.1
	15–30 minutes	99	37.8
	30–60 minutes	23	8.8
	More than 60 minutes	6	2.3
	Homework	157	59.9

Item	Response	n	%
Purposes (select all that apply)	Writing tasks	89	34.0
	Other	73	27.9
	Quizzes/exams	54	20.6
	Writing feedback	46	17.6

The most common frequency was "sometimes" (47.3%). "Sometimes" and "rarely" together accounted for 54.6% of responses and "always" and "often" for 42.4%. Sessions were short, with 88.9% lasting under 30 minutes. Nearly half reported that their use had increased (48.1%), 38.9% reported no change and 13.0% a decrease; this is a retrospective self-report, not a measure of change over time. A majority switched between English and Arabic (57.6%), while 22.9% used English only and 19.5% Arabic only. Homework was the most common purpose (59.9%), followed by writing tasks (34.0%) and quizzes/exams (20.6%).

Focus-group accounts were consistent with occasional, task-triggered use: "There are months when I don't use it"; "I used it maybe once or twice since I joined university"; "I use it only when needed". Use was often tied to specific assessed tasks: "I only use it when I prepare for my speaking presentation"; "I use it only for presentations, with speaking presentations only". Support for comprehension often ran through Arabic: "when I read a text that I don't understand, I ask it to explain it to me or to translate the text into Arabic"; "I sometimes use it to translate texts or audios".

Most participants described no deliberate prompting strategy: "I just write the question in my head". A few described sequenced prompting: "I start with simple question"; "I go step by step ...for example give me the introduction, then give me two body paragraphs, ...give me a conclusion...etc.". In the second example, it appears that prompts are sequenced to have the chatbot generate each section of the essay rather than to refine the student's own text.

6-2- Perceived usefulness for skills and feedback (RQ2)

As Table 3 shows, writing received the highest mean rating (35.5% "highly effective"), followed by speaking and reading. Ratings for writing, speaking and reading were all significantly above the scale midpoint, with small-to-moderate effects ($d = 0.38-0.53$). Listening was rated lowest and its ratings were polarised: 41.2% rated chatbots 1 or 2 for listening and 33.2% rated them 4 or 5, and the mean did not differ from the midpoint. Even under the least favourable pairing of responses, listening was rated lower than writing, $t(261) = 4.47$, speaking, $t(261) = 4.08$, and reading, $t(261) = 3.33$ (all $p \leq .001$, significant after Bonferroni correction).

Table (3)
Perceived Effectiveness of Chatbots For L2 Skills and Overall Perceptions (N = 262)

Item (1–5 scale)	M (SD)	95% CI	% rating 4–5	t(261) vs. 3 (d)
Effectiveness for:				
Writing	3.68 (1.28)	[3.53, 3.84]	60.3	8.61 (0.53)
Speaking	3.63 (1.34)	[3.47, 3.80]	58.0	7.67 (0.47)
Reading	3.51 (1.33)	[3.35, 3.67]	54.2	6.17 (0.38)
Listening	2.95 (1.41)	[2.78, 3.13]	33.2	-0.53 (-0.03), $p = .60$
I agree that...				
My English has improved because of using chatbots	3.73 (0.92)	[3.61, 3.84]	66.8	12.78 (0.79)
I would recommend chatbots to others	3.73 (0.99)	[3.61, 3.85]	63.0	11.92 (0.74)
Chatbots help me learn independently	3.67 (1.00)	[3.55, 3.79]	64.1	10.79 (0.67)

Item (1–5 scale)	M (SD)	95% CI	% rating 4–5	t(261) vs. 3 (d)
I feel more confident using English	3.60 (0.95)	[3.48, 3.71]	57.6	10.22 (0.63)

Note. One-sample t-tests against the scale midpoint of 3; $d = (M - 3)/SD$. All $p < .001$ except where shown. Wilcoxon signed-rank tests led to the same conclusions.

As illustrated in Table 4, students rated chatbots as effective for correcting vocabulary (66.4%) and grammar (61.6%) far more often than for feedback on writing (37.8%) or speaking (35.2%) assignments. Each of these four correction-versus-feedback contrasts remained significant even under the least favourable pairing of responses (McNemar's $\chi^2(1) \geq 13.85$, all $p < .001$). Asked which source they preferred for feedback on their writing, 191 students (72.9%) chose their teacher, 68 (26.0%) a chatbot, and 3 (1.1%) wrote in another answer (e.g., "both"). Among students choosing between the two, the preference for teachers was clear: 73.7%, 95% CI [68.1, 78.7], exact binomial $p < .001$.

Table (4)
Perceived Effectiveness of Chatbot Correction and Feedback (N = 262)

Correction or feedback area	% rating effective
Correcting vocabulary	66.4
Correcting grammar	61.6
Feedback on writing	37.8
Feedback on speaking	35.2

Focus-group participants illustrated these skill-specific uses. For writing: "It helps me a lot with my writing, I ask it to write a passage and then I modify it as I wish"; "after I prepare my writing essay, I ask it to check for any grammatical mistakes or errors in the structure". For reading: "When I use it for reading, I ask it to explain the passage and give me questions about it"; "I ask it to give me a paragraph about a topic, then I read the paragraph and try to understand". For vocabulary: "it is very good especially for vocabulary learning, I ask ChatGPT to give me synonyms for the same word I have, and this helps me increase my vocabulary knowledge". One student distinguished speaking from listening: "I use it to talk to, as there is not enough time to practice speaking in the classroom, I ask chatbot to do a conversation and this improved my speaking skills, but it wasn't effective in developing listening skills". Another noted, "we learn new words from voice conversation and its right pronunciation".

All four overall agreement items were rated well above the midpoint, with moderate-to-large effects ($d = 0.63$ – 0.79 ; see Table 3). Of the students, 66.8% agreed with the statement that their English had improved because of using chatbots (10.3% disagreed), 64.1% that chatbots helped them learn independently (14.1% disagreed), and 57.6% that they felt more confident using English (13.0% disagreed); 63.0% would recommend chatbots to others. It is worth noting that these are self-reported perceptions, not measured learning gains.

6-3- Chatbot use in relation to classroom instruction (RQ3)

Participants positioned chatbots as supplementary to teachers: "it is an aid not a replacement of teachers"; "I use it sometimes as someone help me with understanding something, like I'm explaining to my friend, but it can never be instead of teachers". Interestingly, they reported inconsistent instructor stances: "Some doctors allow it, others don't like us using it" ("doctors" being students' term for university instructors). Several described use that took place outside, rather than within, instruction and wanted that to change: "If I were a teacher, I would use it on the board and explain from it, this will teach us how to use chatbots effectively and will help the teacher too"; "I think if our teachers make it part of their teaching instructions, it will help us a lot.....for example if the teacher presents the task and asks us to use it in front of him/her and we discuss the findings, this will be more productive". Teachers were the first source of chatbot awareness for 25.2% of students, compared with 53.4% for social media, yet teachers were the preferred feedback source (72.9%) and the most common verification resource. As the questionnaire did not measure classroom integration directly, RQ3 rests mainly on the qualitative strand.

6-4- Challenges, verification and support (RQ4)

As shown in Table 5, incorrect information (43.5%) and unclear responses (35.1%) were the most frequently reported challenges, followed by internet connectivity (26.0%) and cost (23.3%). Students described having to repair failed exchanges: "sometimes it gives wrong information or doesn't understand the question itself, so I have to rewrite or explain my question several times". Asking a teacher was the most common verification method, and students described it as their default, turning to other sources when teachers were unavailable: "I cannot trust the answers, I have to check it with my teacher first"; "If my teacher is not available, I usually compare chatbots answers with another resource like google or other chatbot apps like copilot".

Table (5)
Challenges, Verification Methods and Support Needs (Select All That Apply)

Category	Response	n	%
Challenges	Incorrect information	114	43.5
	Unclear responses	92	35.1
	Internet connectivity	68	26.0
	Cost	61	23.3
Output verification	Ask teacher	121	46.2
	Compare with other sources	57	21.8
	Check textbook	54	20.6
	Other (several chatbots, not checking, Google)	30	11.5
Needed support	Teacher guidance	121	46.2
	Training workshops	102	38.9
	Arabic tutorials	84	32.1
	More user-friendly interface	59	22.5

Note. Categories chosen by fewer than 5% of respondents are omitted.

Participants also voiced concern about overreliance: "It became like an addiction.....they rely on it for everything"; "The student no longer knows how to think"; "it is a double-edged-sword...useful but potentially addictive". The support students wanted most was teacher guidance (46.2%), followed by workshops (38.9%) and Arabic tutorials (32.1%), delivered early and by skill: "there must be workshops on how to use it"; "....it can be included as a module in IT classes"; "English teachers can train us on how to use it with each skill.....this should be done at the beginning of each academic year"; "the teacher should train us first, then we can use it".

6-5- Integration of the quantitative and qualitative strands

Table 6 brings the two strands together for each research question. The focus groups mostly confirmed the questionnaire patterns and explained them.

Table (6)
Joint Display of Quantitative and Qualitative Findings

RQ	Questionnaire (N = 262)	Focus groups (n = 15)	Integrated inference
RQ1: Use	ChatGPT 91.2%; smartphone 80.5%; "sometimes" 47.3%; English-Arabic switching 57.6%; sessions under 30 minutes 88.9%; homework 59.9%	Occasional, task-triggered use; translation into Arabic for comprehension; little planned prompting; sequential essay generation	Confirmation and expansion: use is occasional, bilingual and task-driven, and some "strategic" prompting shifts the composing work to the chatbot
RQ2: Usefulness	Writing, speaking and reading above neutral; listening lower than each; correction rated effective more often than feedback; 72.9% prefer teacher feedback	Help with writing, reading and vocabulary; voice practice helped speaking but not listening	Confirmation: chatbots are valued for form-focused support, and the listening gap reflects how students use them
RQ3: L2 Instruction	Teachers are the first source of awareness for 25.2% (social media 53.4%)	Chatbots "an aid not a replacement"; inconsistent instructor stances; wish for classroom demonstration	Expansion: use develops outside instruction, and students ask teachers to guide usage
RQ4: Challenges & support	Incorrect information 43.5%; unclear responses 35.1%; asking a teacher the most common check (46.2%); teacher guidance 46.2%, workshops 38.9%, Arabic tutorials 32.1%	Repairing failed exchanges; distrust of answers; worry about overreliance; requests for early, skill-specific training	Confirmation: verification is deferred to teachers, and students want guided, bilingual training

7- Discussion

The findings are interpreted in light of the conceptual framework. How resources, language and devices mediate students' chatbot use is first examined; following is the consideration of who does the cognitive and linguistic work. Lastly, AI literacies students display or lack, including how they rely on teachers to judge chatbot output are deliberated.

7-1- Occasional, smartphone-based, bilingual use outside instructional design

Students used chatbots occasionally and for specific tasks, mostly in short smartphone sessions, and homework was the most common purpose. This qualifies the picture of near-constant use suggested by surveys of general undergraduate populations (Alshamy et al., 2025). This finding coincides with Ali et al.'s (2025) finding that occasional users use chatbots narrowly, while more frequent use spans more purposes. Three contextual features concerning the resources that mediate use help explain the pattern. First, most students learned about chatbots through social media rather than from teachers. As such, their habits likely formed before and outside any instruction. They were shaped in an online space where learners already circulate views of ChatGPT's benefits for language learning (Chen et al., 2026). Second, dependence on smartphones (80.5%) may lead to short, transactional exchanges over sustained drafting or extended conversation. Third, subscription cost (23.3%) and connectivity (26.0%) barriers are in accordance with Warschauer et al.'s (2023) "rich-get-richer" contradiction; that is, unequal access to reliable connections, devices and possibly paid tiers may translate into unequal learning opportunities.

The rate of Arabic–English switching (57.6%) appears to be the study's most distinctive descriptive finding. Together with accounts of translating texts and audio into Arabic, it suggests that students use their first language as a resource for comprehension. This aligns with the prominence of translation among foundation-year learners (Al-Raimi et al., 2024). From a mediational perspective, L1 use can scaffold access to difficult input, but it can also shift the cognitive work of L2 comprehension to the tool. The data cannot distinguish these possibilities, but this opens a window for process-level research using chat logs. Pedagogically, the demand for Arabic tutorials (32.1%) suggests that AI literacy instruction in foundation programmes should itself be bilingual.

7-2- Who does the work? Prompting, outsourcing and overreliance

Only a few students described any planned prompting, and the most elaborate example ("give me the introduction, then give me two body paragraphs...") is sequential outsourcing of an essay rather than refinement of the student's own writing. Rather than strategic competence, it illustrates Warschauer et al.'s (2023) with-or-without contradiction. To illustrate, the students who most need to develop composing skills can use chatbots to bypass them. Other accounts, such as asking the chatbot to write a passage and then modifying it, or checking one's own essay for errors, leave more of the work with the learner. This variation is what a distributed-agency perspective would predict (Godwin-Jones, 2024). The pedagogical question is therefore less whether students use chatbots than how they divide the work with them.

Students' own references to "addiction" and to students who no longer know "how to think" support review evidence linking overreliance to weaker critical reasoning (Zhai et al., 2024). They also echo experimental evidence that unguided chatbot use can improve task performance while weakening unassisted learning and metacognitive engagement (Bastani et al., 2025; Fan et al., 2025). These accounts show students' concern about this risk, not evidence that it occurs. The use of chatbots for quizzes and exams (20.6%) also warrants institutional attention. Though the item does not distinguish revision from use during assessment, under either reading it underscores the need for explicit assessment policies (Chan & Hu, 2023; Miao & Holmes, 2023).

7-3- Uneven skill affordances and the listening gap

The higher ratings for writing and reading than for listening, which was the only skill rated no differently from neutral, corroborate the research emphasis on writing (Lo et al., 2024) and learners' emphasis on form-focused help (Barrot, 2023; Xiao & Zhi, 2023). The low listening ratings are unlikely to reflect the technological capabilities per se. By 2025 the major chatbots had offered voice interaction, and AI-mediated speaking practice had shown beneficial results (Fathi et al., 2024). A more plausible explanation lies in how students interacted with chatbots: they used them mainly in text, often in Arabic, and for written tasks. This is consistent with evidence that chatbot effects are moderated by modality, interface design and interaction capability (Lyu et al., 2025; Wang et al., 2025). For example, one student commented that voice conversation helped speaking "but it wasn't effective in developing listening skills." This suggests that listening benefits do not follow automatically from spoken interaction without deliberate task design. The polarised ratings fit this: a third of students, perhaps those exploring voice features, rated chatbots useful for listening. Listening therefore appears to be a gap in students' repertoires and in instruction rather than an inherent limitation of the tools.

The gap between perceived accuracy in correcting vocabulary and grammar (61.6–66.4%) and in feedback on writing and speaking assignments (35.2–37.8%) suggests that students trust chatbots more for local, checkable form than for global, task-specific evaluation. This mirrors evidence on the feedback itself. In a pre-departmental intensive English programme comparable to a GFP, ChatGPT's feedback on A1–B1 learners' writing matched teachers' feedback for about two-thirds of grammar and mechanics comments but under a quarter of comments on content and organisation (Cengiz & Yigitoglu Aptoula, 2026). Human feedback

on secondary students' essays was also rated more accurate than ChatGPT's and better at prioritising essential features (Steiss et al., 2024). Plausibly, students see assignment-level feedback as requiring knowledge of the task, rubric and course that only teachers have (Henderson et al., 2026). In terms of AI literacy, students thus display evaluative judgement of local form but defer task-level judgement to teachers. this reliance is examined in the next section.

7-4- Teachers as epistemic anchors

Three findings converge on the central role of teachers: a preference for teacher feedback (72.9%), asking a teacher as the most common verification strategy, and teacher guidance as the most requested form of support. Students treat teachers, not chatbots, as the reference point for judging chatbot outputs. This supports the complementarity view (Jeon & Lee, 2023; Teng, 2024). It also mirrors a survey of almost 7,000 university students, about nine in ten of whom trusted teacher feedback, against six in ten for GenAI feedback (Henderson et al., 2026). However, it complicates evidence that AI feedback can yield gains comparable to human feedback (Escalante et al., 2023). To explain, comparable gains do not imply comparable trust, especially among novice learners in transition. Why students prefer teacher feedback, whether because of alignment with assessment, curricular knowledge or trust, warrants direct investigation.

This reliance on teachers, however, has a drawback. Students who ask a teacher to confirm chatbot answers depend on someone who is usually absent since most chatbot use occurs outside class. More importantly, they do not practise judging the output themselves, although the ability to evaluate AI critically is central to AI literacy (Almatrafi et al., 2024; Long & Magerko, 2020; Miao & Shiohira, 2024; Yang et al., 2025) and yet varies widely among L2 learners (Pan & Zhang, 2026). Some students already take a first step towards independence: when no teacher is available, they compare a chatbot's answer with Google or another chatbot. This strategy is unsystematic and used by only about one in five students (21.8%), but L2 instruction could build on it by teaching students how to check AI output against reliable sources.

7-5- A gap between practice and L2 pedagogy

Students' accounts describe a gap between widespread informal use and formal L2 instruction that is widened by inconsistent instructor stances. AI chatbots seem to have entered students' activity system without corresponding changes to the instruction, tasks and division of labour set by the curriculum (Yang & Kyun, 2022). Students' own proposals, namely teacher demonstration, guided in-class use and discussion of outputs, closely resemble recommendations in the literature (Kohnke et al., 2023; Miao & Holmes, 2023). Evidence from research with trained Omani students (Jahan & Akbar, 2026) and from foundation-year students in Kazakhstan (Aypay & Bedeker, 2025) suggests that this is achievable.

7-6- Implications

For foundation programmes in Oman and the wider Gulf, the findings support four steps. First, English teachers should provide early, skill-specific and bilingual AI literacy instruction covering prompting that keeps the composing work with the learner, independent verification routines, and voice modes for listening and speaking. Second, programme-wide guidelines should replace inconsistent individual stances and specify permitted uses in homework and assessment. Because the national GFP standards, which already include computing and general study skills, are currently under review (Elamin, 2025), AI literacy could also be embedded at national level, as one student proposed: "it can be included as a module in IT classes". Third, task designs should make chatbot use visible, for example through submitted prompt logs and in-class critique of outputs as suggested by critical digital literacies approaches (Darvin, 2025). Lastly, institutions should address device, data-cost and connectivity barriers so that AI-supported tasks do not widen inequalities.

7-7- Contribution

Empirically, the study profiles foundation-year chatbot use in terms of frequency, language choice, device, verification and access barriers. Conceptually, it reads these practices as mediated activity outside instructional design. Teachers are portrayed as epistemic anchors. The AI literacy that would make this mediation productive, particularly the independent evaluation of output, is largely deferred to teachers. This helps explain why positive perceptions coexist with limited strategic use. Contextually, it adds evidence from foundation-year students in Oman to research that has focused mainly on undergraduate and graduate students. Pedagogically, it identifies student-endorsed targets for AI literacy instruction though it does not demonstrate learning effects.

8- Limitations and Future Research

Some critical limitations should be considered when interpreting the findings. Convenience sampling at a single private institution limits transferability. Because participation in the online questionnaire was voluntary (26.2% response rate), students more interested in chatbots may be over-represented. The outcomes are self-reported rather than measured gains. The cross-sectional design cannot capture change beyond one retrospective item. Because teachers recommended articulate students for the focus groups, more engaged or proficient students may be over-represented, and grouping by level may have shaped the discussions. Because members of the research team facilitated the focus groups, students may have given socially desirable accounts of

teachers and approved uses. The study also did not examine cultural factors or specific tool versions, such as free versus paid tiers or voice modes.

Future research should use chat logs or screen recordings to establish who does the work in learner–chatbot interactions. Further investigations may follow cohorts longitudinally across the foundation year and test teacher-mediated AI literacy interventions against comparison groups using objective L2 learning outcome measures. Other studies may investigate the effect of multimodal chatbot use on oracy along with literacy skills.

9- Conclusion

Foundation-year students use GenAI chatbots, mostly ChatGPT on smartphones, often bilingually and mainly for homework and writing. They do so largely outside L2 instruction. They value form-focused support but doubt chatbots' accuracy and therefore rely on teachers to decide what to trust. The central finding is not enthusiasm but a mismatch; that is, students are actively using a powerful tool while the curriculum remains largely silent about it. Interestingly, students themselves are asking teachers to fill that gap. For L2 learning, the implication is that a chatbot's value depends on how the work is divided among learner, tool and teacher. For Omani and Gulf foundation programmes, the foundation year appears to be the appropriate period to teach that division of labour explicitly, early and bilingually before habits of outsourcing become established.

Ethics Statement: The study was approved by the Department of Research in the host institution before data collection. Questionnaire respondents gave electronic informed consent, and focus-group participants gave written consent. Participation was voluntary and had no bearing on students' grades or academic standing. Questionnaire responses were collected anonymously. Data were stored on password-protected devices that are accessible only to the research team.

Declaration of Generative AI Use: AI was used only to identify and format references and edit the language as well as format the manuscript in line with the reporting guidelines of the journal. The authors take full responsibility for the content of this publication.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

ORCID ID: <https://orcid.org/0009-0001-0034-7544>

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

- [1] bdaljaleel, M., Barakat, M., Alsanafi, M., Salim, N. A., Abazid, H., Malaeb, D., Mohammed, A. H., Hassan, B. A. R., Wayyes, A. M., Farhan, S. S., El Khatib, S., Rahal, M., Sahban, A., Abdelaziz, D. H., Mansour, N. O., AlZayer, R., Khalil, R., Fekih-Romdhane, F., Hallit, R., . . . Sallam, M. (2024). A multinational study on the factors influencing university students' attitudes and usage of ChatGPT. *Scientific Reports*, 14, Article 1983. <https://doi.org/10.1038/s41598-024-52549-8>
- [2] Ali, E. A. M., Veettil, R. P., Al-Zaabi, M. A., Sharafudeen, M., & Qureshi, T. (2025). Exploring Omani EFL learners' perceptions of AI tools. *Digital Technologies Research and Applications*, 4(3), 206–216. <https://doi.org/10.54963/dtra.v4i3.1922>
- [3] Al-Issa, A. S., & Al-Bulushi, A. H. (2012). English language teaching reform in Sultanate of Oman: The case of theory and practice disparity. *Educational Research for Policy and Practice*, 11(2), 141–176. <https://doi.org/10.1007/s10671-011-9110-0>
- [4] Almatrafi, O., Johri, A., & Lee, H. (2024). A systematic review of AI literacy conceptualization, constructs, and implementation and assessment efforts (2019–2023). *Computers and Education Open*, 6, Article 100173. <https://doi.org/10.1016/j.caeo.2024.100173>
- [5] Al-Raimi, M., Mudhsh, B. A., Al-Yafaei, Y., & Al-Maashani, S. (2024). Utilizing artificial intelligence tools for improving writing skills: Exploring Omani EFL learners' perspectives. *Forum for Linguistic Studies*, 6(2), Article 1177. <https://doi.org/10.59400/fls.v6i2.1177>
- [6] Al-Rusheidi, I. S. (2021). General Foundation Programs (GFPs) student entry and exit standards. *Education Quarterly Reviews*, 4(4), 382–391. <https://doi.org/10.31014/aior.1993.04.04.400>
- [7] Alshamy, A., Al-Harthi, A. S. A., & Abdullah, S. (2025). Perceptions of generative AI tools in higher education: Insights from students and academics at Sultan Qaboos University. *Education Sciences*, 15(4), Article 501. <https://doi.org/10.3390/educsci15040501>
- [8] Aypay, B., & Bedeker, M. (2025). Exploring the role of ChatGPT in foundation year students' academic literacy practices. In M. Bedeker, T. M. Makoelle, & S. A. Manan (Eds.), *Exploring academic writing as social practice* (pp. 269–294). Palgrave Macmillan. https://doi.org/10.1007/978-3-032-01210-4_12

- [9] Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, 57, Article 100745. <https://doi.org/10.1016/j.asw.2023.100745>
- [10] Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakçı, Ö., & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26), Article e2422633122. <https://doi.org/10.1073/pnas.2422633122>
- [11] Benayoune, A., Slimi, Z., & Al Habsi, A. (2026). Artificial intelligence policy challenges and institutional readiness in Omani higher education. *Discover Education*, 5, Article 86. <https://doi.org/10.1007/s44217-026-01188-4>
- [12] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- [13] Cengiz, Y., & Yigitoglu Aptoula, N. (2026). Comparing AI and human feedback at higher education: Level appropriateness, quality and coverage. *Language Teaching*, 59(1), 122–127. <https://doi.org/10.1017/S0261444825000199>
- [14] Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, Article 43. <https://doi.org/10.1186/s41239-023-00411-8>
- [15] Chen, Z., Wei, W., & Zou, D. (2026). Generative AI technology and language learning: Global language learners' responses to ChatGPT videos in social media. *Interactive Learning Environments*, 34(2), 907–920. <https://doi.org/10.1080/10494820.2025.2511248>
- [16] Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- [17] Crosthwaite, P., & Sun, S. (2026). Generative AI and L2 written feedback studies: A scoping review. *RELC Journal*, 57(1), 207–219. <https://doi.org/10.1177/00336882251386530>
- [18] Darvin, R. (2025). The need for critical digital literacies in generative AI-mediated L2 writing. *Journal of Second Language Writing*, 67, Article 101186. <https://doi.org/10.1016/j.jslw.2025.101186>
- [19] Dehdary, N., & Meschi, R. (2024). The role of ChatGPT in language education: A study of Omani students' perspectives. *Studies in Technology Enhanced Learning*, 4(1). <https://doi.org/10.21428/8c225f6e.6c5360d1>
- [20] DeVellis, R. F., & Thorpe, C. T. (2022). *Scale development: Theory and applications* (5th ed.). SAGE.
- [21] Elamin, M. (2025). Quality audits of the general foundation program in Oman: A review of trends and priorities. *East Journal of Human Science*, 1(5), 206–215. <https://doi.org/10.63496/ejhs.Vol1.Iss5.179>
- [22] Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20, Article 57. <https://doi.org/10.1186/s41239-023-00425-2>
- [23] Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, 56(2), 489–530. <https://doi.org/10.1111/bjet.13544>
- [24] Fathi, J., Rahimi, M., & Derakhshan, A. (2024). Improving EFL learners' speaking skills and willingness to communicate via artificial intelligence-mediated interactions. *System*, 121, Article 103254. <https://doi.org/10.1016/j.system.2024.103254>
- [25] Godwin-Jones, R. (2024). Distributed agency in second language learning and teaching through generative AI. *Language Learning & Technology*, 28(2), 5–30. <https://doi.org/10.64152/10125/73570>
- [26] Henderson, M., Bearman, M., Chung, J., Fawns, T., Buckingham Shum, S., Matthews, K. E., & de Mello Heredia, J. (2026). Comparing generative AI and teacher feedback: Student perceptions of usefulness and trustworthiness. *Assessment & Evaluation in Higher Education*, 51(5), 863–878. <https://doi.org/10.1080/02602938.2025.2502582>
- [27] 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. <https://doi.org/10.1111/jcal.12610>
- [28] Huang, Z., Nazar, N., Abbas, G., & Jiang, Z. (2025). Assessing AI literacy in second language writing: A scale development and validation study. *Education and Information Technologies*, 30(18), 25909–25935. <https://doi.org/10.1007/s10639-025-13767-3>
- [29] Jahan, F., & Akbar, H. (2026). Exploring ChatGPT in EFL classrooms: Insights into engagement, proficiency and ethics. *Discover Education*, 5, Article 174. <https://doi.org/10.1007/s44217-026-01158-w>
- [30] Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 28(12), 15873–15892. <https://doi.org/10.1007/s10639-023-11834-1>
- [31] Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>

- [32] Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research* (5th ed.). SAGE.
- [33] Li, M., Wang, Y., & Yang, X. (2025). Can generative AI chatbots promote second language acquisition? A meta-analysis. *Journal of Computer Assisted Learning*, 41(4), Article e70060. <https://doi.org/10.1111/jcal.70060>
- [34] Lo, C. K., Yu, P. L. H., Xu, S., Ng, D. T. K., & Jong, M. S.-y. (2024). Exploring the application of ChatGPT in ESL/EFL education and related research issues: A systematic review of empirical studies. *Smart Learning Environments*, 11, Article 50. <https://doi.org/10.1186/s40561-024-00342-5>
- [35] Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. <https://doi.org/10.1145/3313831.3376727>
- [36] Lyu, B., Lai, C., & Guo, J. (2025). Effectiveness of chatbots in improving language learning: A meta-analysis of comparative studies. *International Journal of Applied Linguistics*, 35(2), 834–851. <https://doi.org/10.1111/ijal.12668>
- [37] Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://doi.org/10.54675/EWZM9535>
- [38] Miao, F., & Shiohira, K. (2024). *AI competency framework for students*. UNESCO. <https://doi.org/10.54675/JKJB9835>
- [39] Oman Academic Accreditation Authority. (2008). *Oman academic standards for general foundation programs*. <https://oaaqa.gov.om/External-Quality-Assurance/General-Foundation-Programme-Accreditation/>
- [40] Oman Academic Accreditation Authority. (2017). *General foundation program quality audit manual*. <https://oaaqa.gov.om/External-Quality-Assurance/General-Foundation-Programme-Accreditation/>
- [41] Pan, Z., & Zhang, Y. (2026). What does AI literacy mean in L2 writing? A mixed-methods investigation of Chinese EFL student writers' AI literacy profiles and their links with writing self-regulated learning strategies. *International Journal of Applied Linguistics*. Advance online publication. <https://doi.org/10.1111/ijal.70248>
- [42] Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Warschauer, M., & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, Article 101894. <https://doi.org/10.1016/j.learninstruc.2024.101894>
- [43] Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. *Computers and Education: Artificial Intelligence*, 7, Article 100270. <https://doi.org/10.1016/j.caeai.2024.100270>
- [44] Wang, F., Cheung, A. C. K., Neitzel, A. J., & Chai, C. S. (2025). Does chatting with chatbots improve language learning performance? A meta-analysis of chatbot-assisted language learning. *Review of Educational Research*, 95(4), 623–660. <https://doi.org/10.3102/00346543241255621>
- [45] Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second or foreign language. *Journal of Second Language Writing*, 62, Article 101071. <https://doi.org/10.1016/j.jslw.2023.101071>
- [46] Weidlich, J., Gašević, D., Drachsler, H., & Kirschner, P. (2025). ChatGPT in education: An effect in search of a cause. *Journal of Computer Assisted Learning*, 41(5), Article e70105. <https://doi.org/10.1111/jcal.70105>
- [47] Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages*, 8(3), Article 212. <https://doi.org/10.3390/languages8030212>
- [48] 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*, 38(5), 180–210. <https://doi.org/10.14742/ajet.7492>
- [49] Yang, Y., Zhang, Y., Sun, D., He, W., & Wei, Y. (2025). Navigating the landscape of AI literacy education: Insights from a decade of research (2014–2024). *Humanities and Social Sciences Communications*, 12, Article 374. <https://doi.org/10.1057/s41599-025-04583-8>
- [50] Zare, J., Al-Issa, A., & Ranjbaran Madiseh, F. (2025). Interacting with ChatGPT in essay writing: A study of L2 learners' task motivation. *ReCALL*, 37(3), 385–402. <https://doi.org/10.1017/S0958344025000035>
- [51] Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11, Article 28. <https://doi.org/10.1186/s40561-024-00316-7>