Australasian Journal of Educational Technology, 2026, 42(1).

AI-enhanced informal digital learning of English: Effects on EFL students' cognitive, non-cognitive and oral proficiency skills

Amira Desouky Ali
Department of English Language, Faculty of Languages and Translation, Sadat Academy for Management Sciences, Egypt

This study examined the effects of integrating artificial intelligence (AI) tools into informal digital learning of English (IDLE) to enhance cognitive and non-cognitive skills, as well as listening and speaking proficiency among English as a Foreign Language students. A sample of 120 Egyptian university students participated in a mixed-methods design that consisted of a questionnaire, pretests and post-tests for listening and speaking skills and semi-structured interviews. Quantitative data were analysed using descriptive statistics, t tests and mixed analysis of variance, while qualitative responses were thematically explored. The findings revealed significant advancements in cognitive skills, including the regulation of attitudinal needs, goal commitment, resource allocation and metacognitive skills, as well as enhanced non-cognitive skills. However, social connections via AI were found to be less impactful, with many students reporting limited authentic interactions. While AI-driven IDLE significantly enhanced speaking proficiency, listening skills showed more modest gains, suggesting differential effects of AI on productive versus receptive skills. Despite technical challenges, AI-based IDLE demonstrated potential for personalising learning. Future research should address these challenges while focusing on bridging the gap between informal digital learning and real-world language use.

Implications for practice or policy:

- Educators should integrate AI tools into blended learning models, combining AI-driven practice with real-world communicative opportunities to bridge the gap between simulations and authentic language use.
- Developers must prioritise customisation in AI tools, such as adaptive learning paths and realistic conversation practice, to address diverse learner needs effectively.
- Policymakers and administrators should invest in resolving technical barriers (e.g., speech recognition accuracy, Internet reliability) to optimise AI tool effectiveness and user experience.

Keywords: informal digital learning of English (IDLE), cognitive and non-cognitive skills, listening and speaking, AI tools, EFL students

Introduction

Informal digital learning of English (IDLE) has emerged as a prominent subfield within computer-assisted language learning, focusing on self-directed English language acquisition that transcends traditional classroom environments. As defined by Lee (2019a), IDLE refers to the informal and unstructured learning of English through various digital platforms, including mobile devices, computers, social media, websites and online forums. Unlike formal education, IDLE allows learners to select tools and methods, fostering personalised learning experiences. Studies have evidenced that IDLE positively influences language skills, encompassing proficiency and communicative competence, particularly through online chats, discussions and multimedia consumption (Lee, 2019a; Lee & Lee, 2021).

The implementation of IDLE provides significant advantages that benefit EFL learners through cognitive and non-cognitive development. The term cognitive skills encompasses mental processes that comprise goal-setting, problem-solving, attention management, understanding and remembering abilities. Task management relies heavily on these skills, as they support both listening comprehension and speaking proficiency (Anderson, 2020). Non-cognitive skills encompass motivational qualities, including enjoyment, which serve as fundamental elements for learners to persist and adapt within their educational settings

(Kautz et al., 2014). Through authentic social interaction, the method helps learners build confidence and develop positive attitudes towards language use (Pérez-Paredes, 2024). IDLE serves as an effective addition to standard language instruction by supporting the growth of both cognitive abilities and non-cognitive participation.

IDLE roots its theoretical framework in Zimmerman's (2000) self-regulated learning (SRL) theory and Vygotsky's (1978) sociocultural theory, which together explain learner interaction with IDLE environments enhanced by artificial intelligence (AI). The core components of IDLE's learner-driven approach derive from SRL, which focuses on students taking responsibility for goal setting, strategy selection and reflection on their learning progress. Vygotsky's sociocultural theory explains how social interactions, along with tool mediation systems such as AI, function as cognitive tools that drive language development. Dörnyei's (2009) second language motivational self system, along with socio-emotional learning theory (Kautz et al., 2014), serves as a framework to analyse non-cognitive skills (e.g., motivation, emotional engagement) through examining learners' ideal second language selves and vision, as well as emotional regulation in learning. The combination of these theoretical frameworks demonstrates how AI-based informal learning of English enables students to develop independent language learning abilities alongside intellectual growth and motivational involvement in English in English as a Foreign Language (EFL) contexts.

Recent AI developments have expanded the scope of IDLE to encompass the autonomous use of digital technologies for language practice. This includes AI-powered activities like using speech recognition for pronunciation feedback or conversing with an AI chatbot, a practice known as AI-enhanced IDLE. Through generative AI technology, learners can access customised educational programmes that include natural language processing functions and interactive dialogue systems for skill development (Cai, 2023). Although data privacy concerns exist along with worries about human interaction in education (Chomsky, 2023), AI demonstrates a valuable capacity to support informal language learning.

This study conceptualised AI within IDLE as a dynamic, multi-role tool, guided by the synthesis of sociocultural theory and SRL. AI functions as a tutor by providing scaffolded feedback on pronunciation and grammar, a collaborator by enabling conversational practice and a content generator by personalising learning materials. This fluidity allows learners to exercise agency, using AI to support their specific self-regulatory processes – from goal-setting to reflection – in a fully autonomous environment. In this study, AI did not replace formal instruction but served as a complementary practice space that extended learning beyond the classroom.

Research has firmly established the significance of IDLE in fostering language acquisition, primarily focusing on vocabulary outcomes and motivational factors in global contexts (Lai & Gu, 2011; Lee & Lee, 2021; Soyoof et al., 2023). Concurrently, the broader literature on AI in education has effectively mapped the macro-landscape, charting its use for profiling and prediction (Chu et al., 2022), outlining research agendas for generative AI (Lodge et al., 2023) and conceptualising AI's affordances for developing 21st-century skills (Celik et al., 2024). However, a critical gap exists at the intersection of these fields: a lack of empirical research investigating how specific AI tools are integrated within self-directed IDLE contexts to impact discrete language skills. This study addresses this gap by moving beyond conceptual frameworks and attitudinal surveys to provide mixed-methods evidence from an underrepresented Middle Eastern EFL setting. While prior IDLE work prioritised vocabulary and general motivation, this research specifically demonstrates the actual efficacy and limitations of AI in enhancing listening and speaking proficiency within a formal tertiary education programme. The study was integrated into a listening and speaking course for English majors. Hence, it specifically contributes to understanding how AI can be leveraged to support and extend formal learning outcomes in higher education, offering a model for the purposeful integration of educational technology into university language programmes.

Literature review

IDLE and AI technologies

IDLE encompasses self-directed activities in informal digital contexts, often devoid of teacher assessment, with personal interests driving engagement (Lee & Lee, 2021). Lee (2019b) has distinguished between extramural and extracurricular IDLE activities. While extramural IDLE is purely autonomous, extracurricular IDLE incorporates varying levels of structured teacher involvement.
IDLE for language acquisition involves two main areas: receptive activities, where students consume English-language media, and productive activities, where students interact on social media platforms and create content (Sundqvist & Sylvén, 2016). According to Benson's (2011) explanatory theory, IDLE comprises four fundamental components: the formality level, classroom environment (inside or outside the classroom), learning method (taught, self-taught or natural acquisition) and locus of control (self-driven or other-driven). The different aspects emphasise the evolving and varied characteristics of IDLE.

AI-powered tools, including chatbots and voice assistants, introduce a new dimension to IDLE learning. For instance, voice assistants such as Google Home and Amazon Alexa facilitate real-time conversational practice, exposing learners to diverse accents and speech patterns. Furthermore, intelligent tutors incorporated within language learning applications offer personalised feedback, thus enhancing the overall learning experience (Dizon & Tang, 2020). Additionally, generative AI technologies, such as OpenAI's ChatGPT and Google's Bard, have revolutionised IDLE learning by simulating real-world language interactions and providing personalised, meaningful conversations (Bozkurt & Sharma, 2023). AI chatbot technologies are built on large language models like LaMDA and GPT, which are extensively trained on vast data sets (Kasneci et al., 2023). These AI systems generate contextually appropriate texts, recognise language nuances and manage cultural and genre-specific references with high linguistic accuracy (Liu et al., 2024; Rudolph et al., 2023). Despite promising advancements, studies indicate that challenges remain, such as the potential for AI to generate inaccurate information or students' discomfort when interacting with chatbots (Crompton et al., 2024; Ericsson et al., 2023).

AI-enhanced IDLE effects on cognition and non-cognition

IDLE has demonstrated substantial potential in enhancing cognitive and non-cognitive skills (Lai et al., 2017; Lee & Drajati, 2019). Research evidence has demonstrated that the cognitive abilities of students benefit from improved speaking skills along with enhanced grammar and vocabulary learning (Cole & Vanderplank, 2016; Ericsson et al., 2023). Through promoting learner autonomy, the IDLE system allows students to select resources based on their goals, which results in better language processing and critical thinking capabilities. The interactive multimedia-rich environment of IDLE motivates students to practice language consistently, which leads to improved non-cognitive skills, including motivation, engagement and enjoyment (Lee & Drajati, 2019). According to research, AI tools help students develop metacognition and goal-setting abilities because they require students to define their language learning targets and track their progress. The research conducted by Lai et al. (2024) evaluated SRL together with individual interest to understand self-directed technology-based informal learning activities and discovered that self-regulation serves as a key predictor for cognitively demanding activities. Through social networking platforms and digital resources integrated with AI, students can self-monitor their progress to observe improvements in their pronunciation, fluency and overall communication abilities (Fauziah & Diana, 2023). AI-based speech recognition systems help learners to evaluate their pronunciation while enabling them to make improvements that lead to better language accuracy and metacognitive development (Klimova, 2020; Shivakumar et al., 2019).

This adaptive feedback bolsters learners' capacity to organise and regulate their practices, thus leading to more effective listening comprehension and speaking fluency (Madhavi et al., 2023).

On the non-cognitive front, AI tools have been shown to enhance motivation and confidence in practising language skills (Pérez-Paredes, 2024). This is particularly relevant in overcoming speaking anxiety and low self-confidence through personalised and non-judgemental feedback provided by AI tools (Waloyo et al., 2021). IDLE activities are associated with psychological aspects of second language acquisition, including anxiety (Lee, 2019b), motivation (Fauziah & Diana, 2023; Lee & Drajati, 2019) and confidence and enjoyment (Lai et al., 2015; Lee, 2019a).

In this study, cognitive skills include regulating attitudinal needs, goal commitment, metacognition, resource needs and social connection to help learners focus and engage with AI tools for language learning. Non-cognitive skills encompass confidence and enjoyment to drive learners' ongoing engagement in speaking and listening practices. The selection of cognitive and non-cognitive skills in this study was guided by their theoretical and practical relevance to IDLE and the use of AI tools. From a theoretical perspective, these skills align with Zimmerman's (2000) SRL, Vygotsky's (1978) social constructivism and Dörnyei's (2009) motivational theory, which emphasise the importance of autonomy, interaction and intrinsic motivation in informal learning contexts. Practically, these skills address the real-world challenges learners face when engaging with AI tools in unstructured environments. For example, regulating attitudinal needs and goal commitment helps learners to sustain motivation and focus, while metacognition enables them to adapt their learning strategies based on AI-generated feedback. Social connection fosters authentic communication and emotional support, and resource needs ensure efficient management of time and materials. Non-cognitive skills, such as confidence and enjoyment, drive ongoing engagement and persistence, making them essential for successful informal learning.

Using AI in IDLE listening and speaking

IDLE activities significantly enhance learners' vocabulary, fluency, pronunciation and communicative competence through digital tools such as AI chatbots and social networks. For instance, Lee and Jeon (2022) and Zou et al. (2018) have highlighted that AI chatbots can mimic real-life interactions, thus improving oral proficiency by presenting learners with opportunities to practice their speaking skills. Moreover, integrating social networks with language applications elevates fluency and organisational skills, while auditory and visual aids enhance vocabulary retention (Pérez-Paredes, 2024).

Atmojo (2021) has demonstrated that regular participation in multimodal IDLE activities through audio, lyrics and visuals enhances Indonesian learners' EFL proficiency. Nugroho and Triana (2021) have found that although students maintain positive attitudes towards IDLE, they fail to use it in non-classroom settings because they prefer local languages and face digital device constraints. My study also examined how AI-based IDLE affects learners' ability to communicate effectively. According to Nugroho et al. (2022), receptive and productive IDLE activities helped Chinese EFL learners develop their communicative competence through using AI chatbots such as ChatGPT and Bing Chat. Research has shown that informal digital activities such as watching YouTube videos and engaging with mobile apps lead to improvements in pronunciation, fluency and overall language competence, especially in informal language settings (Ali, 2022; Klimova 2020; Lutfiana et al., 2021; Trinder 2017). EFL learners encounter specific issues when they use AI tools since their English practice remains limited to classroom time (Kazu & Kuvvetli 2023; Madhavi et al., 2023), but these tools might resolve this problem by giving students more practice opportunities.
Although studies to date have outlined IDLE study in Asian and European regions, they lack sufficient investigation into AI-based language learning tools for Middle Eastern countries, including Egypt. Addressing this gap is critical, as the Middle Eastern EFL context – characterised by distinct transfer challenges from first to second language (Alfaifi & Saleem, 2024) and a sociocultural landscape that differs markedly from previously studied regions (Bashori et al., 2020) – provides a crucial test case for the global applicability and cultural transferability of AI-driven language learning solutions.

The research field also shows insufficient coverage of cognitive and non-cognitive processes that take place during listening and speaking activities in IDLE settings. AI applications for fluency and pronunciation development received attention from Lee and Jeon (2022) and Zou et al. (2018). Nevertheless, their studies failed to explore the cognitive mechanisms behind language skill development, including metacognitive awareness, goal commitment and resource utilisation. Research on AI tools controlling non-cognitive elements in IDLE settings shows limited availability. Consequently, this study aimed to address the following research questions (RQs):

1. To what extent does using AI tools outside classrooms affect students' cognitive and non-cognitive skills?
2. To what extent does using AI tools outside classrooms affect students' listening and speaking skills?
3. To what extent do students engage with specific AI tools and activities outside of classrooms to enhance their listening and speaking abilities?
4. What challenges do students encounter while using AI tools?

Methods

Participants

The study was conducted with a sample of 120 students enrolled in an English listening and speaking course within the English major programme at the Sadat Academy, Faculty of Languages and Translation, Egypt. The research population consisted of 120 students, with 90 females (75%) and 30 males (25%), all of whom self-identified as belonging to the middle socio-economic class. The participants consisted of native Arabic speakers with a mean age of 20.4 years (SD = 0.9), ranging from 19 to 22. A stratified random sampling method was employed to ensure a representative and balanced sample. The research sample was divided into three groups according to students' previous AI tool usage patterns: high, medium and low usage. The research collected data on AI tool usage through a pre-study survey administered to all potential participants. The survey included questions about the frequency and duration of AI tool usage. The participants were distributed across usage strata according to their survey responses:

- high usage: participants who reported using AI tools for more than 5 hours per week
- medium usage: participants who reported using AI tools for 2–5 hours per week
- low usage: participants who reported using AI tools for fewer than 2 hours per week.

Once the participants were stratified, random selection was conducted within each category to assign individuals to the experimental or control group. This dual strategy of randomisation and stratification serves two critical purposes: randomisation mitigates the potential for selection bias, while stratification ensures that both experimental and control groups are balanced concerning their prior usage of AI tools. The resulting classification indicated that 52 students were allocated to the control group, while 68 students were assigned to the experimental group. The pre-study survey and stratification process ensured that the groups were comparable in terms of their prior exposure to AI tools, thereby enhancing the validity of the study's findings.

Research design and data collection

This study employed an explanatory sequential mixed-methods design, characterised by an initial quantitative phase followed by a qualitative phase to provide deeper insights. The study commenced with quantitative data collection through pretests and post-tests measuring listening and speaking proficiency, alongside a questionnaire assessing cognitive and non-cognitive factors. This study received ethical approval from the institution and was conducted in accordance with ethical principles, including voluntary informed consent, participant anonymity and the right to withdraw. The quantitative findings directly informed the subsequent qualitative data collected from semi-structured interviews in two key ways:

1. Survey responses (e.g., outliers, unexpected trends) and test score patterns (e.g., high vs low performers) were analysed to identify participants for semi-structured interviews, ensuring representation of diverse experiences.
2. Interview questions were tailored to explore quantitative results in depth.

The study followed this procedural timeline throughout a semester:

- Week 1: pretests (listening and peaking) and baseline IDLE questionnaire administered to both control and experimental groups
- Week 11: post-tests and full IDLE questionnaire for the two groups
- Week 12: semi-structured interviews (purposively sampled participants from the experimental group, stratified by performance and engagement levels).

Integration of data occurred at two stages:

1. During the qualitative phase, interview protocols were adjusted based on emergent survey and test trends.
2. During interpretation, quantitative results were juxtaposed with qualitative themes (e.g., learners attributing progress to AI's instant feedback) to develop a comprehensive understanding of AI-enhanced IDLE's impact. This triangulation strengthened validity and provided nuanced explanations for observed outcomes. The qualitative findings provided explanatory depth to the quantitative results, revealing learners' perceptions about why and how certain outcomes (e.g., skill gains or motivation changes) occurred in their AI-enhanced IDLE experiences.

Instruments

Three validated instruments were used: a questionnaire, pretest and post-test listening and speaking tests and semi-structured interviews.

IDLE questionnaire

The questionnaire, adapted from An et al. (2020) and Lai and Gu (2011), was constructed to assess the utilisation of AI tools in listening and speaking within IDLE environments. It encompasses 34 questions segmented into two sections and was administered online via Google Forms. Part A consists of three domains: demographic information, AI tool usage and activities. This segment gathers data on the participants' frequency of AI tool usage, rated on a 5-point Likert scale from *never* to *always*. Part B focuses on six domains: regulating attitudinal needs, goal commitment, metacognition, resource needs, social connection and non-cognitive skills. These domains were rated from 1 *strongly disagree* to 5 *strongly agree*, assessing how AI tools influence students' motivation, goal-setting, learning strategies, social engagement, and overall confidence and enjoyment in learning English outside classrooms through AI. Test-retest reliability was applied, yielding a Pearson correlation coefficient of 0.80, indicating high reliability.

Listening and speaking tests

To evaluate the impact of IDLE on students' listening and speaking skills, IELTS-based pretests and post-tests were employed to ensure reliable and valid outcomes. These tests were administered in two equivalent forms to ensure comparability. A rater and I undertook the scoring of the tests to guarantee objectivity. The listening test comprised four progressively difficult segments, featuring conversations and monologues across various social and academic contexts, including questions of different types, assessing comprehension of main ideas, details, opinions, structure of spoken texts and attitudes, as well as the ability to follow the development of arguments and paraphrasing. The speaking test was conducted as a face-to-face interview and included three components: general questions, a brief presentation on a specific topic and an in-depth discussion. The evaluation criteria for speaking encompassed content, fluency, lexical resource, grammatical range and pronunciation, with both of us ensuring robust reliability and consistency in scoring. Both tests demonstrated high reliability, evidenced by a Cronbach's alpha of 0.92 for the listening test and an inter-rater reliability score of 0.93 for the speaking test, indicating strong consistency in results and scoring.

Semi-structured interviews

From the experimental group, 15 participants were selected for interviews through stratified purposeful sampling based on (a) pre-posttest improvement quartiles (top 25%, middle 50% and bottom 25% of gain scores), (b) questionnaire responses indicating high (> 4.5), medium (3–4.5), or low (< 3) motivation levels on a 5-point scale and (c) reported frequency of AI tool use (daily, weekly or occasional users). This approach ensured that the interviewees' quantitative profiles represented the full range of experiences with the AI-enhanced IDLE environment. Participants signed an informed consent document outlining the interview's purpose and procedures. A rater and I carried out the interviews, which adhered to a standardised protocol based on guidelines from Jacob and Furgerson (2012) to ensure consistency and minimise bias. Topics of discussion encompassed cognitive and non-cognitive facets, as well as the challenges of using AI tools. A pilot test was utilised to refine the protocol, and probing techniques were employed to elicit detailed responses. Data analysis was conducted with transparency to ensure that confidentiality and anonymity were maintained throughout the study.

AI tool selection and classification

The tools in this study were classified as AI driven based on their use of at least one of the following computational features, as defined by literature on AI in language learning (Klimova, 2020; Luckin et al., 2016):

- real-time natural language processing (NLP) feedback (e.g., pronunciation scoring, syntactic error detection)
- adaptive algorithms (e.g., personalised exercise difficulty based on learner performance)
- generative or dialogic AI (e.g., conversational agents, automated speech synthesis).

Tools lacking these features (e.g., static listening exercises without AI feedback) were excluded from final analysis to ensure alignment with the study's focus on AI-enhanced IDLE. The selected tools fall into five categories, each meeting the above criteria and targeting specific language skills:

- AI speech recognition and feedback tools
  - Google Assistant, Speechling: Provide *real-time pronunciation analysis* using NLP to detect errors in stress, intonation and phonemes.
  - Elsa Speak: uses *AI-powered speech models* to benchmark learners' pronunciation against native speakers.
- AI-powered adaptive learning apps
  - Duolingo, Babbel: employ *adaptive algorithms* to adjust vocabulary and grammar exercises based on learner proficiency (Lee & Jeon, 2022).
- AI-mediated conversational platforms
  - Tandem, HelloTalk: include *AI-driven features* such as translation bots, error-detection algorithms, and chatbot partners (e.g., HelloTalk's AI Chat).
- AI pronunciation and fluency tools
  - Replika: uses *generative AI* to simulate natural conversations and provide contextual feedback.
- AI-enhanced listening tools
  - BBC Learning English (AI features): includes *interactive transcripts* and *automated comprehension checks* (e.g., chatbot mode).
  - FluentU: uses *NLP-driven subtitles* and vocabulary highlighters tailored to learner levels.

All tools are openly available, with most offering free or freemium versions to ensure accessibility for all participants. To address the inherent variability of self-directed AI tool use and ensure a consistent core learning experience across all participants, the following protocol was implemented:

- Structured activity framework: While learners had the autonomy to choose when to engage, they were provided with a structured weekly framework of suggested activities. This included:
  - Minimum engagement: A requirement of 2 hours per week of total AI tool usage.
  - Balanced skill focus: Guidelines to split time evenly between speaking-focused practice (e.g., using Speechling, Replika) and listening-focused practice (e.g., using FluentU, BBC Learning English).
  - Task variety: A menu of specific, repeatable tasks for each tool category (e.g., "Complete one Babbel story lesson", "Engage in a 5-minute conversation with Replika on a suggested topic").
- Orientation and training: All participants attended a mandatory orientation session where I demonstrated core functionalities for each tool, ensuring everyone started with a baseline understanding of how to access and use the key features that provided feedback and adaptive content.
- Monitoring and compliance: Consistency was verified through a mixed-methods approach:
  - Quantitative: Where possible, in-app analytics were used to track session frequency and duration.
  - Qualitative: Participants maintained simple self-report logs to record the tools used and time spent, which were cross-referenced for compliance with the minimum time requirement.

This protocol balanced the informal nature of IDLE with the methodological rigour required for a comparative study, ensuring that all participants received comparable opportunities for input, interactive practice and AI-generated feedback, even as the specific content was adapted to their individual performance.

The flexible implementation of the AI tools – where learners exercised choice within a structured framework – was a deliberate and ecologically valid feature of the study's design. While this introduced a degree of variability, it was essential for two key reasons:

1. To preserve the informal nature of IDLE: The design embraced this by allowing participants to gravitate towards tools and activities they found most engaging, thereby mirroring real-world informal learning and increasing the external validity of the findings.
2. To investigate authentic tool use: The primary research aim was to understand the impact of AI-enhanced IDLE as it is naturally practised, not under artificial laboratory conditions. The observed variability in tool preference and usage patterns is itself a valuable finding, revealing how different learners leverage AI to meet their personal needs.

Procedure

Participants in the experimental group were provided with detailed guidelines on how to use the tools, including a minimum usage requirement of 2 hours per week. They were also given an orientation session to familiarise them with the tools' features and functionalities. To balance methodological rigour with IDLE's informal ethos, participants received suggested activity types and frequencies but retained full autonomy over when or how to engage. To monitor compliance and account for variations in usage, participants were asked to maintain self-reported logs of their activities, and app analytics were used where available. This approach ensured that all participants had equal opportunities to engage with the tools while allowing for individual differences in usage patterns. To ensure that the amount of input was equal across groups, the control and experimental groups were instructed to spend 2 hours per week on out-of-class learning activities. The control group engaged in traditional methods, such as listening to English podcasts and watching English videos, while the experimental group used AI tools for informal learning. Study logs were used to track and verify the time spent on these activities. A detailed framework outlining the study's intervention structure, including tool usage guidelines and activity suggestions for both experimental and control groups, is provided in Appendix A (https://drive.google.com/file/d/1dwbo1oDmG5fQFa5CSF_X-Kd9QgRsAp4_/view?usp=sharing).

Data analysis

Quantitative data were analysed using IBM SPSS Statistics (version 24) to compute descriptive statistics, paired-samples *t* tests, and mixed analysis of variance (ANOVA). Complementing these analyses, qualitative interview transcripts underwent inductive thematic analysis following Braun and Clarke's (2006) framework. Two coders manually conducted iterative coding to ensure reliability: initial open coding identified emergent concepts, followed by axial coding to synthesise these into coherent themes. To strengthen trustworthiness, multiple verification strategies were employed. First, 20% of transcripts were double-coded, achieving strong intercoder reliability (κ = .82). Second, member-checking procedures ensured participants validated interview summaries for accuracy. Finally, methodological triangulation cross-referenced qualitative themes with quantitative questionnaire responses and proficiency test scores, reinforcing the robustness of interpretations.

Results

This section synthesises the multi-method findings, with research questions addressed through statistically significant results and supporting participant narratives.

RQ1: Effect on cognitive and non-cognitive skills

To evaluate the effectiveness of the intervention, a two-way mixed ANOVA was conducted, comparing the experimental and control groups on cognitive and non-cognitive skills before and after the intervention. At baseline, the experimental group (M = 3.06, SD = .35) and the control group (M = 2.89, SD = .92) were equivalent. In the post-intervention, the experimental group reached a perfect maximum score (M = 5.00, SD = .00), while controls improved more modestly (M = 3.33, SD = 1.12), as shown in Table 1.

Table 1
Descriptive statistics of cognitive and non-cognitive skills

Group	Pretest ($M \pm SD$)	Post-test ($M \pm SD$)	Gain (Δ)	Cohen's d
Control	2.89 ± 0.92	3.33 ± 1.12	+0.44	0.24
Experimental	3.06 ± 0.35	5.00 ± 0.00	+1.94	1.49

Cohen's d effect sizes revealed negligible baseline differences between groups (d = 0.24). Following the intervention, the experimental group demonstrated dramatically superior outcomes with very large effects for both post-test scores (d = 1.49) and gain scores (d = 1.94), indicating the treatment's substantial impact. Table 2 presents the mixed ANOVA results of the main effect of time, time × group interaction, and the main effect of group.

Table 2
Mixed ANOVA results of cognitive and non-cognitive skills

Effect	F	p	Partial η^2	Observed power
Within-subjects				
Time	45.903	< .001	.178	1.00
Time × Group	6.430	.012	.029	0.71
Between-subjects				
Group	37.470	< .001	.150	1.00
Intercept	1342.826	< .001	.864	1.00

A significant improvement occurred across both groups over time (F = 45.90, p < .001, η^2 = .178), with a medium-to-large effect size, explaining 17.8% of variance. Greenhouse-Geisser correction (ε = 1.00) confirmed robustness to sphericity violations (Mauchly's p < .001). Additionally, the experimental group improved disproportionately more than controls (F = 6.43, p = .012, η^2 = .029), though this interaction effect was small (2.9% variance explained). The experimental group outperformed controls overall (F =

37.47, p < .001, η^2 = .150), with a medium effect size (15% variance explained). The intercept was significant (F = 1342.83, p < .001), confirming model fit.

To evaluate the differences among students in the experimental group across various domains, a paired-sample t test and descriptive statistics were conducted, with the results presented in Figure 1.

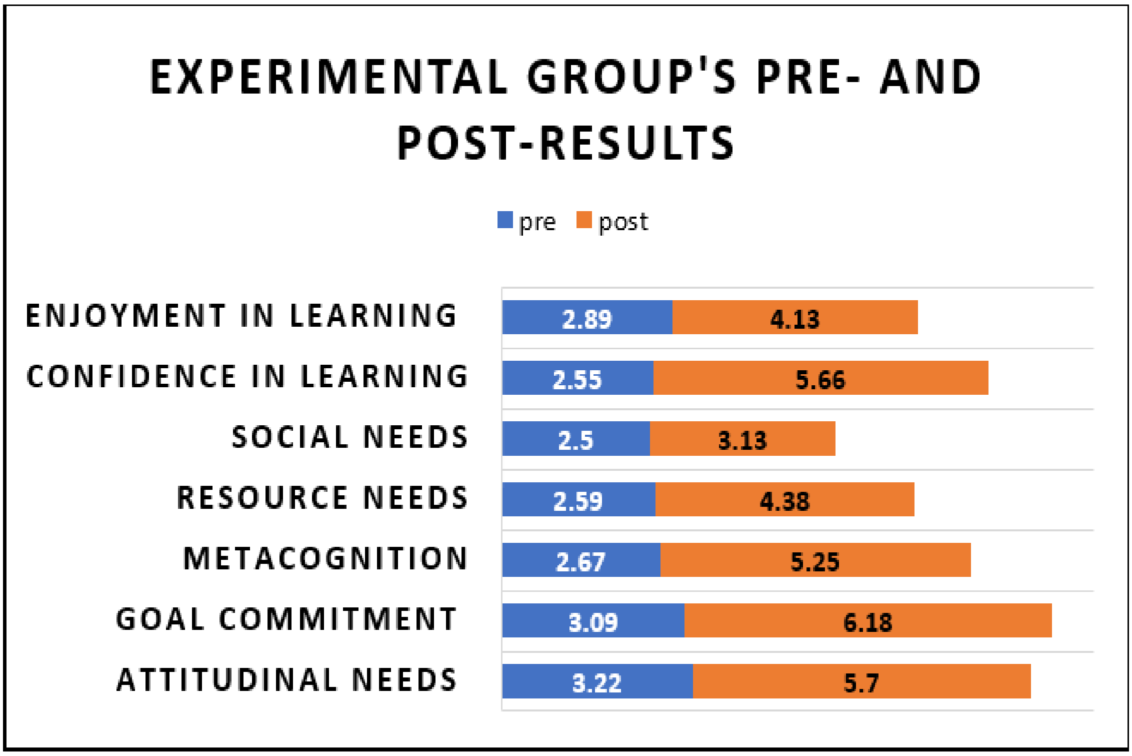

Figure 1. Descriptive statistics of experimental groups' cognitive and non-cognitive skills

The analysis revealed significant improvements in most domains following the utilisation of AI tools. Specifically, attitudinal needs demonstrated a considerable increase, with a pretest mean of 3.22 rising to a post-test mean of 5.7 (t = 5.058, p < .05). Goal commitment also exhibited a significant enhancement, with means increasing from 3.09 to 6.18 (t = 4.009, p < .05). Metacognition showed a noteworthy gain, reflected in the mean increase from 2.67 to 5.25 (t = 6.704, p < .05). Similarly, resource needs experienced a significant rise, with means going from 2.59 to 4.38 (t = 3.204, p < .05). However, social connection did not reveal a statistically significant change, with pre-and post-test means of 2.5 and 3.13, respectively (t = 3.201, p > .05). Lastly, confidence and enjoyment of learning demonstrated considerable improvements, with respective increases from 2.55 to 5.66 (t = 3.054, p < .05) and from 2.89 to 4.13 (t = 6.367, p < .05).

To substantiate these findings with qualitative data derived from interviews, a table summarising the themes and corresponding participant quotations is presented in Appendix B (https://drive.google.com/file/d/1dwbo1oDmG5fQFa5CSF_X-Kd9QgRsAp4_/view?usp=sharing). Below is a brief overview of the results, focusing on both cognitive and non-cognitive skills.

Motivation and interest in learning

AI tools' influence on motivation varied: 40% of interviewees (6/15) reported increased motivation due to engaging content, while another 40% (6/15) noted sustained interest in new topics. However, 20% (3/15) observed no motivational change, maintaining pre-study attitudes towards learning.

Personalisation of learning experience

Perceptions of personalisation diverged: 47% (7/15) viewed AI tools as highly adaptive to their pace and needs, whereas 40% (6/15) found them moderately personalised (somewhat generic). The remaining 13% (2/15) reported minimal personalisation, citing uniform content.

Support in goal-setting and achievement

A total of 80% (12/15) found AI tools effective for goal-setting through structured guidance, while 20% (3/15) felt they offered only practice support without goal-oriented scaffolding.

Access to learning resources

A total of 87% (13/15) highlighted expanded access to global materials, contrasting with 13% (2/15) who saw no significant resource increase.

Social connections

A total of 40% (6/15) reported improved peer/native-speaker interactions, while 53% (8/15) deemed social features inauthentic. One participant (7%) abstained from social functionalities.

Confidence in English

A total of 67% (10/15) experienced major confidence gains from AI practice, 27% (4/15) noted moderate improvements (with lingering anxiety and 13% (2/15) saw minimal change.

Enjoyment in learning

A total of 47% (7/15) enjoyed gamified tools, 33% (5/15) appreciated interest-aligned content and 20% (3/15) found AI tools no more enjoyable than traditional methods.

RQ2: Effect on listening and speaking skills

Participants' oral proficiency skills were assessed using standardised pre- and post-tests, with results reported as mean scores (*M*) ± SD. Table 3 highlights baseline equivalence between the groups at pretest and quantifies skill gains post-intervention, providing insight into the intervention's effectiveness. Regarding listening skills, the groups were nearly equivalent at pretest (control: $M = 25.83 \pm 5.22$; experimental: $M = 26.24 \pm 4.76$), indicating comparable starting proficiency. As for speaking skills, the control group ($M = 27.72 \pm 4.89$) scored slightly in the pretest higher than the experimental group ($M = 25.52 \pm 4.78$), though the difference was modest (Δ = 2.20 points).

Table 3
Descriptive statistics of listening and speaking skills

Skill	Group	Pretest (*M* ± *SD*)	Post-test (*M* ± *SD*)	Gain (Δ)	Cohen's *d*
Listening	Control	25.83 ± 5.22	35.01 ± 9.05	+9.18	1.24
	Experimental	26.24 ± 4.76	37.94 ± 8.97	+11.70	1.52
Speaking	Control	27.72 ± 4.89	26.71 ± 4.65	−1.01	−0.21
	Experimental	25.52 ± 4.78	42.65 ± 5.31	+17.13	3.58

As for the listening post-test, both groups improved, but the experimental group gained more (+11.70 vs +9.18 in controls). In the post-speaking test, the control group regressed slightly (1.01 points), while the experimental group improved dramatically (+17.13 points), far surpassing the control group's post-test score (42.65 vs 26.71). This suggests the intervention had a large, positive effect on speaking proficiency.

To rigorously evaluate the intervention's effectiveness on speaking and listening skills, a two-way mixed ANOVA was conducted, examining the effects of time (pretest vs post-test), group (experimental vs control) and their interaction – an approach that accounts for within-subject development and between-group differences simultaneously. Table 4 displays the mixed ANOVA results of listening and speaking skills.

Table 4
Mixed ANOVA results of listening and speaking skills

Source	*df*	*F*	*p*	Partial η²	Observed power
Listening					
Time	1,118	356.24	< .001	.465	1.00
Time × Group	1,118	5.17	.023	.012	.61
Group	1,118	10.55	.002	.025	.89
Speaking					
Time	1,118	618.12	<. 001	.601	1.00
Time × Group	1,118	783.50	<. 001	.656	1.00
Group	1,118	293.86	<. 001	.417	1.00

For listening skills, results showed more nuanced patterns. A large main effect of time ($F[1,118] = 356.24$, $p <. 001$, $\eta^2 = .465$) indicated significant improvement across both groups. However, the time × group interaction was only marginally significant ($F[1,118] = 5.17$, $p = .023$, $\eta^2 = .012$), suggesting the intervention provided a slight additional benefit beyond natural improvement. The small effect size (1.2% variance explained) and suboptimal power (.61) for this interaction imply limited confidence in its practical significance. A modest but significant main effect of group ($F[1,118] = 10.55$, $p = .002$, $\eta^2 = .025$) revealed the experimental group maintained slightly better listening scores overall, though this accounted for only 2.5% of variance.

The mixed ANOVA revealed significant effects across all tested dimensions for speaking skills. A massive main effect of time was observed ($F[1,118] = 618.12$, $p <. 001$, $\eta^2 = .601$), indicating both groups showed dramatic improvement from pre- to post-testing. The exceptionally large time × group interaction ($F[1,118] = 783.50$, $p < .001$, $\eta^2 = .656$) demonstrated that the experimental group's gains far exceeded those of controls, confirming the intervention's effectiveness. Additionally, a substantial main effect of group ($F[1,118] = 293.86$, $p < .001$, $\eta^2 = .417$) showed the experimental group maintained superior overall performance. All speaking effects reached perfect observed power (1.00), indicating the analyses were optimally sensitive to detect these robust effects.

While the intervention produced overwhelming effects on speaking skills (with effect sizes exceeding $\eta^2 = .40$), its impact on listening was markedly smaller. The dramatic speaking improvements ($\eta^2 = .656$ for interaction) contrasted with the minimal listening benefits ($\eta^2 = .012$), suggesting the intervention's mechanisms may have been more potent for productive than receptive skills. Power analysis confirmed the study was well equipped to detect speaking effects (1.00) but underpowered for subtle listening interactions (.61).

To assess the impact of the AI-based IDLE intervention on the experimental group's listening and speaking subskills, a paired-samples *t* test was conducted comparing pretest and post-test performance.

Table 5
Paired-sample t test of listening skills on post-test

Skills	Tests	*Means*	*SD*	*T*	*df*	Sig. (2-tailed)	Cohen's *d*
Main idea	Pretest	6.16	1.663	-10.025	67	< .001	0.94
	Post-test	8.00	2.204				
Specific information	Pretest	6.43	2.191	-6.348	67	< .001	0.62
	Post-test	7.78	2.109				
Opinion	Pretest	6.32	2.150	-9.387	67	< .001	0.92
	Post-test	8.28	2.066				
Argument	Pretest	6.46	1.691	-9.879	67	< .001	0.90
	Post-test	8.19	2.115				
Details	Pretest	6.30	1.309	-12.518	67	< .001	1.12
	Post-test	8.28	2.129				
Paraphrase	Pretest	6.80	1.811	-7.614	67	< .001	0.84
	Post-test	8.35	2.049				
Structure	Pretest	6.00	1.574	-10.501	67	< .001	0.96
	Post-test	7.86	2.238				
Total	Pretest	24.78	4.717	-26.111	67	< .001	2.52
	Post-test	42.13	8.477				

Results in Table 5 indicates a significant increase in the experimental group's total mean scores of the listening skills, with a mean difference of 17.35, (SD = 3.76, t = -26.111, and p < .001). Notable improvements were observed in the mean differences of these subskills: listening for the main idea (1.84), specific information (1.35), opinion (1.96), argument (1.73), details (1.98), paraphrasing (1.55) and structure (1.86). Cohen's *d* values ranged from 0.62 to 1.12, indicating a substantial positive impact of the intervention on listening skills.

These improvements were reinforced by qualitative data highlighting how AI tools expanded accessible learning opportunities. As one participant noted, "With AI tools, I could listen to news and conversations anytime – even while commuting. The more I listened to real speech, the easier it became to catch key points", This finding suggests that exposure to diverse, on-demand listening materials through AI platforms enhanced both comprehension breadth (main ideas) and depth (details).

Table 6 demonstrates the results of the paired-sample t test on speaking skills within the experimental group before and after the intervention.

Table 6
Paired-sample t test of speaking skills in post-test

Skills	Tests	*Means*	*SD*	*T*	*df*	Sig. (2-tailed)	Cohen's *d*
Content	Pretest	4.30	2.519	-9.800	67	< .001	0.62
	Post-test	5.55	1.239				
Vocabulary	Pretest	3.73	2.830	-12.180	67	< .001	0.92
	Post-test	5.88	1.701				
Pronunciation	Pretest	3.67	2.483	-10.555	67	< .001	0.93
	Post-test	5.68	1.775				
Accuracy	Pretest	3.95	2.180	-9.589	67	< .001	1.05
	Post-test	5.90	1.436				
Fluency	Pretest	3.79	2.430	-11.527	67	< .001	1.08
	Post-test	5.83	1.093				
Total	Pretest	27.76	4.874	-30.448	67	< .001	2.78
	Post-test	42.37	5.601				

The results in Table 6 show a significant increase in mean scores, with a mean difference of 14.61, (SD = 0.727, t = -30.448, and p < .001). Substantial growth was noted in the mean differences of the subskills:

content (1.25), vocabulary (2.15), pronunciation (2.01), accuracy (1.95), and fluency (2.04). Cohen's *d* values ranged from 0.62 to 1.08, indicating substantial improvement across all subskills. These results reveal the effectiveness of the intervention in enhancing speaking proficiency.

Interview responses showed the same pattern, with users emphasising how real-time AI feedback boosted their confidence and skills. One participant noted, "At first I hesitated to speak, but Speechling fixed my pronunciation the moment I finished a sentence. Watching my accuracy score climb week after week pushed me to practise more". Another participant highlighted fluency gains: "The chat drills on Replika felt friendly, so I no longer freeze mid-sentence". These results indicate that personalised, low-pressure AI interactions encouraged more practice and incremental skill development.

RQ3: Frequency of using AI tools and activities

Before the intervention, both experimental and control groups reported minimal engagement with AI tools, with fewer than 5% of participants using them regularly. Following the intervention, the experimental group demonstrated a 60% increase in AI tool adoption. Post-intervention usage patterns revealed that 40% of participants used AI tools *always*, 30% *often*, 20% *sometimes* and 10% *rarely*, with no respondents selecting *never*. Time investment varied, with 45% of learners dedicating 1–2 hours daily, 23% spending 3–4 hours, 15% engaging for 5–6 hours and 17% reporting less than 1 hour of use.

Three primary AI tool categories contributed to these outcomes, each associated with distinct learning gains. Speech recognition and pronunciation tools, adopted by 35% of participants (e.g., Google Assistant, Speechling), led to a 22% improvement in pronunciation accuracy compared to the control group. This enhancement was facilitated by real-time NLP feedback that identified and corrected phoneme-level errors. Adaptive learning platforms (e.g., Duolingo, Babbel), used by 60% of learners, correlated with an 18% increase in listening comprehension scores, as their algorithmically adjusted exercises matched individual proficiency levels while maintaining optimal cognitive load. The most widely adopted tools were conversational AI systems (e.g., Replika, Elsa Speak), with 83% usage; these contributed to a 2.3-fold increase in fluency markers, such as reduced pauses, particularly among the 23% of learners who engaged for 3–4 hours daily.

Notably, FluentU, an AI-powered listening tool preferred by 71% of participants, was associated with a 31% reduction in self-reported listening anxiety, likely due to its interactive transcription features. In contrast, infrequent users (10% of the sample) showed 40% smaller gains, underscoring the dose-dependent nature of AI tool efficacy.
The experimental group's progress was driven by three key system-level advantages of AI technologies: personalised learning through adaptive algorithms, immediate feedback on pronunciation and grammar and immersive conversational practice available at any time. Together, these features created a tailored and flexible learning environment that supported skill development.

Participants’ responses in the interviews directly reflected these advantages of AI tools. Many highlighted personalisation, with one noting, "Duolingo adapted to my level, focusing on my weak areas like a tutor". Instant feedback was frequently mentioned, as in,"Speechling’s real-time scoring showed me exactly which sounds to fix". Others emphasised immersive practice, stating, "Replika let me converse anytime without fear of judgement". These accounts demonstrate how AI’s technical capabilities fostered consistent, targeted, and anxiety-free practice.

RQ4: Technical and experiential challenges

Ten interviewees encountered various technical challenges, with difficulties concerning understanding different accents and unstable internet connectivity being the most prevalent issues. Additionally, three participants pointed to the lack of natural, spontaneous interaction as a significant drawback, noting that the AI tools failed to replicate the dynamic nature of real-life conversations. For two participants, an initial steep learning curve associated with using the tools further hindered their engagement, requiring additional time and effort to become comfortable with the interface and functionality.

Qualitative comments about technical hurdles matched quantitative results: 38% of the experimental group scored at least ≥ 20% lower on tasks needing real-time AI interaction (e.g., speech recognition exercises) than they did on offline activities. This drop appeared tied to connectivity problems. Members of this group also logged 30% fewer practice sessions each week (M = 3.2 compared to 4.6) than their peers who did not report such barriers.

Discussion

This study contributes to the growing body of literature on AI-enhanced language learning by providing empirical evidence regarding the impact of specific AI tools on listening and speaking skills within the context of IDLE. This research specifically investigated the nuanced ways in which AI tools influence both cognitive and non-cognitive skills. By integrating qualitative and quantitative data, the study presents a comprehensive perspective on learners' experiences, while addressing underexplored challenges. Furthermore, the findings broaden current discussions by offering insights into the specific tools and activities learners engage with, providing valuable guidance for educators and AI developers seeking to enhance language learning through technology.

The quantitative outcomes of the research indicate that tool-specific effects exist: learners using speech recognition tools (Google Assistant and Speechling) reported 22% greater pronunciation accuracy. Recent research from Huang et al. (2023) has evidenced that these tools, particularly NLP tools such as Speechling, far outperform traditional approaches to learning segmental phonemes. This corroborates with Knight et al. (2023), where tools with visible feedback mechanisms (e.g., waveform displays in Speechling) foster metacognitive awareness. AI’s cognitive benefits were clearest in speaking proficiency, where paired-sample t tests showed significant improvements linked to tools providing immediate feedback (e.g., Replika’s conversational practice). Similarly, Jegede's (2024) findings extended this result by revealing that fluency gains were dose dependent: students using AI tools 3–4 hours daily showed 2.3 times more improvement than occasional users. Qualitative data contextualised this by noting that AI’s 24/7 availability enables deliberate practice – a key advantage over classroom constraints, as reported by Haines (2025). However, these improvements coexisted with what Zawacki-Richter et al. (2019) termed the *black box problem*: 62% of interviewed learners could not explain why Replika corrected certain phrases and 40% accepted feedback uncritically. This aligns with Kayalı et al.'s (2023) findings, which confirmed ChatGPT's dual nature as a user-friendly tool that provides fast, relevant responses while also posing risks related to performance errors and the provision of misleading feedback. This underscores the need to pair AI tools with explicit critical AI literacy training, enabling learners to interrogate feedback mechanisms rather than passively internalising them.

While speaking skills showed dramatic AI-driven gains, listening improvements were more nuanced. Though both groups progressed in listening, the experimental group’s additional +2.52 gain over the control group was marginally significant (p = .023, η^2 = .012)—suggesting AI tools like Duolingo and Babbel provided only slight advantages beyond natural development. This aligns with Zawacki-Richter et al.’s (2019) caution that listening comprehension, reliant on implicit acquisition and external exposure, may resist isolated AI interventions. The modest group effect (η^2 = .025) further implies that tool design should better target bottom-up listening processes (e.g., phoneme discrimination), not just top-down practice.

Non-cognitive outcomes exhibited more variation. While 75% of participants reported heightened confidence, qualitative responses revealed that tool efficacy was differentiated: pronunciation apps (Elsa Speak) boosted self-efficacy through granular feedback, whereas conversational AI (Replika) reduced anxiety by simulating low-stakes interactions. However, 25% of learners found AI interactions overly scripted, echoing Godwin-Jones’ (2019) caution about generic implementations. This tool-specific nuance explains the mixed experiences in personalisation and underscores the need for more adaptive algorithms.

The lack of social connection gains emerged consistently across methods. Quantitative surveys showed no significant improvement, while interviews attributed this to AI’s current inability to replicate human spontaneity – a finding that triangulates with Kurian's (2023) and Liu et al.'s (2024) work identifying empathy gaps in conversational AI. This finding illustrates Zawacki-Richter et al.'s (2019) depersonalisation limitation, where AI's transactional nature fails to foster genuine social-emotional learning. Hence, AI should augment rather than replace human interaction in pedagogy.

Additionally, key design insights emerged from tool engagement patterns. The short-term nature of the intervention raises the possibility that the observed gains may partially reflect a novelty effect – where initial enthusiasm for AI tools temporarily boosted engagement. Other studies (Garcia et al., 2025; Jeno et al., 2019) have cautioned that novelty-driven motivation often stabilises or declines after 3–4 months of tool use. While the qualitative data showed sustained engagement during the study period, longitudinal research is needed to determine whether these improvements persist beyond the intervention timeframe or adapt to learners' evolving expectations. Moreover, speech recognition tools exhibited 35% uptake but recorded the highest satisfaction scores across all tools (4.2/5). The focused functionality – immediate and socially acknowledged – meant that learner needs were addressed. While the multifunctional platform FluentU achieved high adoption (71%), its satisfaction ratings were comparatively lower (3.5/5). Qualitative data highlighted interface complexity as a key deterrent, with learners reporting difficulty navigating its multiple features. This aligns with Liu et al.'s (2022) finding that diverse learning approaches yield divergent software satisfaction levels. Therefore, FluentU's all-in-one design – though popular – may overwhelm users seeking specialised tasks. The quantitative and qualitative discrepancies noted above suggest a trade-off between possibility or versatility and usability, which aligns with Twabu's (2025) cognitive load principles for multimodal AI.

A couple of unexpected findings emerged from integrated analysis: (a) Intermediate learners benefited most from AI (40% greater gains than beginners), likely due to sufficient baseline skills to utilise feedback effectively. (b) Even infrequent AI users (10% of sample) outperformed the control group, suggesting minimal thresholds for efficacy. Although technical limitations (connection problems, recognition errors) were mentioned during analysis, the study´s sequential design revealed they had a regional hazard – the qualitative data illustrated that technical issues disproportionately affected technologically constrained learners (30% of the sample) in that limited bandwidth and older devices compromised the use of the AI tools as a reliable resource, which explains the bimodal distribution of the quantitative sample.

Limitations and suggestions for future research

This study's findings are tempered by several limitations. A key risk in postgraduate contexts is overreliance on AI, which may inhibit independent critical thinking and problem-solving. While AI efficiently corrects surface errors, overdependence can bypass the cognitive struggle vital for deep learning. Concerns about academic integrity and homogenised, impersonal output also arise if students misuse generative AI.

The study's sample from Egyptian EFL learners limits generalisability to other populations. Self-reported data may include response bias, and uneven participant tool usage means some AI functions are over- or under-represented. Technical issues like connectivity and speech recognition inaccuracies further highlight that AI implementation remains imperfect.

Future research should employ longitudinal studies to assess if short-term gains in fluency and motivation are sustained. Comparative work could analyse AI's impact on discrete skills (e.g., reading versus speaking), and cross-cultural studies would test consistency across diverse learners. Most pressingly, research must investigate hybrid models combining AI scalability with human interaction to bridge the current authenticity gap in conversation practice.

Implications and conclusion

For educators and developers, AI tools require greater customisation for individual learner needs, especially in adaptive learning and realistic conversation. I recommend a blended approach, integrating

AI practice with real-world communication, to narrow the gap between simulation and authentic use. Developers must enhance speech recognition, connectivity and opportunities for authentic interaction.

In conclusion, this study contributes an evidence-based evaluation of AI in language education. AI can democratise practice, offering accessible, personalised learning that builds confidence. However, it is not a universal solution. Tools must be pedagogically driven, serving as supplements to—not replacements for—human-led instruction. This evidence guides the innovative and ethical use of technology, acknowledging its potential while respecting its limitations to build better, fairer language learning opportunities.

References

Alfaifi, A., & Saleem, M. (2024). Negative transfer and delay in proficiency development: L1 Influenced syntax issues faced by Arab EFL learners. *Forum for Linguistic Studies*, *6*(4), 42–57. https://doi.org/10.30564/fls.v6i4.6664

Ali, A. (2022). Effects of a gamified MALL application on developing EFL preparatory school students' speaking and enjoyment, and reducing speaking anxiety. *CDELT Occasional Papers in the Development of English Education*, *78*(1), 241–296. https://doi.org/10.21608/opde.2022.249943

An, Z., Gan, Z., Wang, C. (2020). Profiling Chinese EFL students' technology-based self-regulated English learning strategies. *PLoS One, 15*(10), Article e0240094. https://doi.org/10.1371/journal.pone.0240094

Anderson, J. R. (2020). *Cognitive psychology and its implications* (9th ed.). Worth Publishers.

Atmojo, A.E.P. (2021). Roles of informal digital learning of English in developing an EFL teacher's English proficiency. *Research and Innovation in Language Learning 4*(3), 266–283. https://doi.org/10.33603/rill.v4i3.5685

Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2020). Web-based language learning and speaking anxiety. *Computer Assisted Language Learning*, *35*(5-6), 1058–1089. https://doi.org/10.1080/09588221.2020.1770293

Benson, P. (2011). Language learning and teaching beyond the classroom: an introduction to the field. In P. Benson & H. Reinders (Eds), *Beyond the language classroom: The theory and practice of informed language learning and teaching* (pp. 7–16). Palgrave Macmillan.

Bozkurt, A., & Sharma, R. C. (2023). Challenging the status quo and exploring the new boundaries in the age of algorithms: Reimagining the role of generative AI in distance education and online learning. *Asian Journal of Distance Education, 18*(1), i-viii. https://www.asianjde.com/ojs/index.php/AsianJDE/article/view/714/397

Braun, V., & Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology,* 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cai, W. (2023, February 24). *ChatGPT can be powerful tool for language learning*. University Affairs. https://www.universityaffairs.ca/career-advice/career-advice-article/chatgpt-can-be-powerful-tool-for-language-learning/

Celik, I., Gedrimiene, E., Siklander, S., & Muukkonen, H. (2024). The affordances of artificial intelligence-based tools for supporting 21st-century skills: A systematic review of empirical research in higher education. *Australasian Journal of Educational Technology*, *40*(3), 19–38. https://doi.org/10.14742/ajet.9069

Chomsky, N. (2023). *The false promise of ChatGPT*. The New York Times. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

Chu, H.-C., Hwang, G.-H., Tu, Y.-F., & Yang, K.-H. (2022). Roles and research trends of artificial intelligence in higher education: A systematic review of the top 50 most-cited articles. *Australasian Journal of Educational Technology*, *38*(3), 22–42. https://doi.org/10.14742/ajet.7526

Cole, J., & Vanderplank, R. (2016). Comparing autonomous and class-based learners in Brazil: Evidence for the present-day advantages of informal, out-of-class learning. *System, 61*, 31–42. https://doi.org/10.1016/j.system.2016.07.007

Crompton, H., Edmett, A., Ichaporia, N., & Burke, D. (2024). AI and English language teaching: Affordances and challenges. *British Journal of Educational Technology, 55*(2), 1–27. https://doi.org/10.1111/bjet.13460

Dizon, G., & Tang, D. (2020). Intelligent personal assistants for autonomous second language learning: An investigation of Alexa. *JALT CALL Journal, 16*(2), 107–120. https://doi.org/10.29140/jaltcall.v16n2.273

Dörnyei, Z., & Muir C. (2019). Creating a motivating classroom environment. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 719–736). Springer. https://doi.org/10.1007/978-3-030-02899-2_36

Ericsson, E., Hashemi, S. S., & Lundin, J. (2023). Fun and frustrating: Pupils' perspectives on practicing speaking English with virtual humans. *Cogent Education, 10*(1), Article 2170088. https://doi.org/10.1080/2331186X.2023.2170088

Fauziah, F., & Diana, N. (2023). Exploring students' Informal Digital Learning of English (IDLE) and self-regulated language learning from a sociocultural perspective. *Indonesian TESOL Journal, 5*(2), 197–214. https://doi:10.24256/itj.v5i2.4225

Garcia, M. B., Goi, C. L., Shively, K., Maher, D., Rosak-Szyrocka, J., Happonen, A., Bozkurt, A., & Damaševičius, R. (2025). Understanding student engagement in AI-powered online learning platforms: A narrative review of key theories and models. In A. Gierhart (Ed.), *Cases on enhancing P-16 student engagement with digital technologies* (pp. 1–30). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3693-5633-3.ch001

Godwin-Jones, R. (2019). Riding the digital wilds: Learner autonomy and informal language learning. *Language Learning & Technology, 23*(1), 8–25. https://doi.org/10.64152/10125/44667

Haines, H. K. (2025). Upskilling entrepreneurship education: Deliberate practice standards, cognitive task analysis, and harnessing AI for skill development. *Entrepreneurship Education and Pedagogy*. https://doi.org/10.1177/25151274251337752

Huang, A. Y., Lu, O. H., and Yang, S. J. (2023). Effects of artificial intelligence–enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Compututers & Education*, *194*, Article 104684. https://doi.org/10.1016/j.compedu.2022.104684

Jacob, S. A., & Furgerson, S. P. (2012). Writing interview protocols and conducting interviews: Tips for students new to the field of qualitative research. *Qualitative Report,* 17(42), 1–10. https://doi.org/10.46743/2160-3715/2012.1718

Jegede, O. (2024). Artificial Intelligence and English language learning: Exploring the roles of AI-driven tools in personalizing learning and providing instant feedback. *Universal Library of Languages and Literatures, 1*(2), 6–19. https://doi.org/10.70315/uloap.ullli.2024.0102002

Jeno, L. M., Vandvik, V., Eliassen, S., & Grytnes, J.-A. (2019). Testing the novelty effect of an m-learning tool on internalization and achievement: A self-determination theory approach. *Computers & Education, 128,* 398–413. https://doi.org/10.1016/j.compedu.2018.10.008

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., and Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success* (OECD Education Working Papers, No. 110). OECD Publishing. https://doi.org/10.1787/5jxsr7vr78f7-en

Kayalı, B., Yavuz, M., Balat, Şener, & Çalışan, M. (2023). Investigation of student experiences with ChatGPT-supported online learning applications in higher education. *Australasian Journal of Educational Technology*, *39*(5), 20–39. https://doi.org/10.14742/ajet.8915

Kazu, İ. Y. & Kuvvetli, M. (2023). The influence of pronunciation education via artificial intelligence technology on vocabulary acquisition in learning English. International *Journal of Psychology and Educational Studies, 10*(2), 480–493. https://doi.org/10.52380/ijpes.2023.10.2.1044

Klimova, B. (2020). Benefits of the use of mobile applications for learning a foreign language by elderly population. *Procedia Computer Science, 176*, 2184–2191. https://doi.org/10.1016/j.procs.2020.09.255

Knight, S., Dickson-Deane, C., Heggart, K., Kitto, K., Çetindamar Kozanoğlu, D., Maher, D., Narayan, B., & Zarrabi, F. (2023). Generative AI in the Australian education system: An open data set of stakeholder recommendations and emerging analysis from a public inquiry. *Australasian Journal of Educational Technology*, *39*(5), 101–124. https://doi.org/10.14742/ajet.8922

Kurian, N. (2023). AI's empathy gap: The risks of conversational Artificial Intelligence for young children's well-being and key ethical considerations for early childhood education and care. *Contemporary Issues in Early Childhood*, *26*(1), 132–139. https://doi.org/10.1177/14639491231206004

Lai, C., Chen, Q., Wang, Y., & Qi, X. (2024). Individual interest, self-regulation, and self-directed language learning with technology beyond the classroom. *British Journal of Educational Technology, 55*(1), 379–397. https://doi.org/10.1111/bjet.13366

Lai, C., & Gu, M. (2011). Self-regulated out-of-class language learning with technology. *Computer Assisted Language Learning, 24* (4), 317–335. https://doi.org/10.1080/09588221.2011.568417

Lai, C., Hu, X., & Lyu, B. (2017). Understanding the nature of learners' out-of-class language learning experience with technology. *Computer Assisted Language Learning*, *31*(1–2), 114–143. https://doi.org/10.1080/09588221.2017.1391293

Lai, C., Zhu, W., & Gong, G. (2015). Understanding the quality of out-of-class English learning. *TESOL Quarterly, 49*(2), 278–308.https://doi.org/10.1002/tesq.171

Lee, J. S. (2019a). Informal Digital Learning of English and second language vocabulary outcomes: Can quantity conquer quality? *British Journal of Educational Technology, 50*(2), 767–778. https://doi.org/10.1111/bjet.12599

Lee, J. S. (2019b). Quantity and diversity of informal digital learning of English. *Language Learning & Technology, 23*(1), 114–126. https://doi.org/10125/44675

Lee, J. S., & Drajati, N. A. (2019). Affective variables and informal digital learning of English: Keys to willingness to communicate in a second language. *Australasian Journal of Educational Technology, 35*(5), 168–182. https://doi.org/10.14742/ajet.5177

Lee, J. S., & Jeon, J. (2022). Visualizing a disembodied agent: Young EFL learners' perceptions of voice-controlled conversational agents as language partners. *Computer Assisted Language Learning, 37*(5–6), 1048–1073. https://doi.org/10.1080/09588221.2022.2067182

Lee, J. S., & Lee, K. (2021). The role of informal digital learning of English and L2 motivational self-system in foreign language enjoyment. *British Journal of Educational Technology, 52*(1), 358–373. https://doi.org/10.1111/bjet.12955

Liu, G., Darvin, R., & Ma, C. (2024): Exploring AI-mediated informal digital learning of English (AI-IDLE): a mixed-method investigation of Chinese EFL learners' AI adoption and experiences. *Computer-Assisted Language Learning*. https://doi.org/10.1080/09588221.2024.2310288

Lodge, J. M., Thompson, K., & Corrin, L. (2023). Mapping out a research agenda for generative artificial intelligence in tertiary education. *Australasian Journal of Educational Technology*, *39*(1), 1–8. https://doi.org/10.14742/ajet.8695

Lutfiana, Z., Kristina, D., & Sumardi, M. (2021). Online Informal Learning of English activities to improve students' English communication skills. *Al-Ishlah: Jurnal Pendidikan, 13*(3), 1909–1916. https://doi.org/10.35445/alishlah.v13i3.1018

Madhavi, E., Sivapurapu, L., Koppula, V., Rani, E., & Sreehari, V. (2023). Developing learners' English-speaking skills using ICT and AI tools. *Journal of Advanced Research in Applied Sciences and Engineering Technology, 32*(2), 142–153. https://doi.org/10.37934/araset.32.2.142153

Nugroho, A., Hong Van, D., Putra, H. &, Nanda, G. (2022). To what extent does Informal Digital Learning of English contribute to learners' communicative competence? *Tadris: Jurnal Keguruan dan Ilmu Tarbiyah 7*(2), 409–422. https://doi.org/10.24042/tadris.v7i2.13500

Nugroho, A., & Triana, Y. (2021). EFL learners' beliefs and practices on Informal Digital Learning of English beyond classroom. *IJEE (Indonesian Journal of English Education), 8*(2), 198-212. https://doi.org/10.15408/ijee.v8i2.19843

Pérez-Paredes, P. (2024) Data-driven learning in informal contexts? Embracing broad data-driven learning (BDDL) research. In P. Crosthwaite (Ed.), *Corpora for language learning: Bridging the research-practice divide* (pp. 211–226). Routledge. https://doi.org/10.4324/9781003413301-15

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching, 6*(1), 1–22. https://doi.org/10.37074/jalt.2023.6.1.9

Shivakumar, A., Shukla, S., Vasoya, M., Kasrani, I. M., & Pei, Y. (2019). AI-enabled language speaking coaching for dual language learners. *IADIS International Journal on WWW/Internet, 17*(1), 66–78. https://www.iadisportal.org/ijwi/papers/2019171105.pdf

Soyoof, A., B. L. Reynolds, B. Vazquez-Calvo, & K. McLay. (2023). Informal digital learning of English (IDLE): A scoping review of what has been done and a look towards what is to come. *Computer Assisted Language Learning, 36*(4), 608–640. https://doi.org/10.1080/09588221.2021.1936562.

Sundqvist, P., & Sylvén, L. K. (2016). *Extramural English in teaching and learning: From theory and research to practice*. Springer. https://doi.org/10.1057/978-1-137-46048-6

Trinder, R. (2017). Informal and deliberate learning with new technologies. *ELT Journal, 71*(4), 401–412. https://doi.org/10.1093/elt/ccw117

Twabu, K. (2025). Enhancing the cognitive load theory and multimedia learning framework with AI insight. *Discover Education, 4*, Article 160. https://doi.org/10.1007/s44217-025-00592-6

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Waloyo, A., Khoiriyah., & Farah, R. (2021). Teachers' perception to CLIL and web-based material implementation in a primary school. *English Review: Journal of English Education, 9*(2), 227–234. https://doi.org/10.25134/erjee.v9i2.4347

Zawacki-Richter, O., Marín, V.I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators?. *International Journal of Educational Technology in Higher Education*, *16,* Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*(1), 82–91. https://doi.org/10.1006/ceps.1999.1016

Zou, D., Xie, H., & Wang, F. L. (2018). Future trends and research issues of technology-enhanced language learning: A technological perspective. *Knowledge Management & E-Learning, 10*(4), 426–440. https://doi.org/10.34105/j.kmel.2018.10.026

Corresponding author: Amira Desouky Ali, amira.ali@sadatacademy.edu.eg

Please cite as: Ali, A. D. (2026). AI-enhanced informal digital learning of English: Effects on EFL students' cognitive, non-cognitive and oral proficiency skills. *Australasian Journal of Educational Technology*, *42*(1), 18–37. https://doi.org/10.14742/ajet.10506