



Journal of International Students
Volume 16, Issue 20 (2026), pp. 1-18
ISSN: 2162-3104 (Print), 2166-3750 (Online)
jistudents.org
<https://doi.org/10.32674/t6wvv124>

Role of Artificial Intelligence in Transforming Adult English Language Learning: A Mixed-Methods Experimental Study

Süleyman Kasap,
Department of English Language Teaching, Van Yüzüncü Yıl University, Turkey
Email: suleymankasap@yyu.edu.tr
ORCID: 0000-0001-8367-8789

ABSTRACT: The rapid proliferation of artificial intelligence (AI) has created new affordances for second language acquisition (SLA), particularly in adult education. However, empirical evidence regarding the efficacy of AI-integrated instruction for adult learners facing unique challenges such as time constraints and foreign language anxiety (FLA) remains lacking. This study employed an explanatory sequential mixed-methods design to investigate the impact of AI-integrated pedagogy on English proficiency and anxiety levels. The sample consisted of 72 adult English as a foreign language (EFL) learners ($N=72$) divided into an experimental group ($n=36$) utilizing generative AI tools (e.g., chatbots, voice assistants) and a control group ($n=36$) following a traditional textbook-based curriculum over a 16-week period. Quantitative analysis revealed that compared with the control group, the experimental group achieved statistically significant gains in overall proficiency ($p < .001$), particularly in speaking and writing. Furthermore, the AI group demonstrated a drastic reduction in foreign language anxiety levels. The qualitative findings from the semistructured interviews indicated that the learners perceived the AI as a “nonjudgmental” practice partner that provided immediate feedback and enhanced learner autonomy. These results suggest that AI integration offers a scalable solution for overcoming affective and logistical barriers in adult language education.

Keywords: Artificial Intelligence in Education, Adult Learning, Foreign Language Anxiety, EFL, Mixed Methods, Generative AI.

Received: March 16, 2026 | **Revised:** June 19, 2026 | **Accepted:** July 15, 2026

How to Cite (APA): Kasap, S. (2026). The Role of Artificial Intelligence in Transforming Adult English Language Learning: A Mixed-Methods Experimental Study. *Journal of International Students*, 16(20), 1-32.

INTRODUCTION

Education is now undergoing major change, largely because of the use of artificial intelligence (AI). Over the last twenty years, classroom technology has moved from basic computer-based drills to flexible, responsive systems that can mimic real human interaction (Chen et al., 2022). This shift is especially clear in language learning, where the focus is moving away from simply memorizing grammar rules and toward building real communication skills within rich, digital environments (Kinsemolu & Onyeaka, 2025).

For adult learners, acquiring a second language (L2) presents distinct challenges compared with younger students. Adults often struggle with “foreign language anxiety” (FLA), a psychological barrier that inhibits speech production because of the fear of negative evaluation by peers or instructors (Bakkar, 2019; El Shazly, 2021). Furthermore, adult learners frequently face time constraints that limit their access to traditional tutoring, necessitating more flexible, self-regulated learning solutions (Wei, 2023).

Recent literature suggests that AI tools—specifically large language models (LLMs) and voice recognition software—can address these specific pain points. By providing a “safe,” nonjudgmental environment for practice and

offering immediate, personalized feedback, AI has the potential to lower affective filters and accelerate proficiency (Huang et al., 2023).

Despite the theoretical promise of AI, there is a need for rigorous empirical validation of its effectiveness in specific contexts. While bibliometric analyses indicate a surge in research interest regarding AI in education (Liu et al., 2025; Borgohain et al., 2024), many existing studies rely on short-term interventions or focus primarily on university-aged students in academic settings. There is a paucity of experimental research focusing specifically on general adult learners in nonuniversity contexts over sustained periods.

Furthermore, while it has been established that AI can assist with translation or writing correction (Deng & Yu, 2022; Li et al., 2024), less is known about the psychological impact of these tools. Does interacting with a machine actually reduce the anxiety associated with language learning, or does it merely provide a mechanical drill? Understanding the relationships among AI integration, anxiety reduction, and academic success is critical for designing effective adult education curricula.

This study addresses these gaps by implementing a 16-week quasiexperimental intervention with 72 adult learners. By comparing a traditional instruction model with an AI-integrated blended learning model, this research aims to provide concrete evidence of “meaningful success”. The findings contribute to the fields of applied linguistics and educational technology by demonstrating how AI can be leveraged not to replace human instructors but to increase their capacity to teach anxious or time-constrained adult learners (Kundu & Bej, 2025).

To guide this investigation, the study poses the following three research questions (RQs), aligning with the mixed-methods design:

- **RQ1 (Quantitative):** Is there a statistically significant difference in English proficiency scores between adult learners taught using an AI-integrated method (Experimental Group) and those taught using a traditional method (Control Group) after a 16-week period?
- **RQ2 (Quantitative):** To what extent does AI-integrated instruction affect the foreign language anxiety (FLA) levels of adult learners compared with traditional instruction?
- **RQ3 (Qualitative):** How do adult learners perceive the role of AI tools in their learning process, particularly in terms of their confidence and motivation?

Literature Review

The integration of artificial intelligence (AI) into educational settings has shifted from a theoretical possibility to a practical necessity. Over the past two decades, the landscape of education has transformed significantly, with AI playing a central role in redefining how knowledge is delivered and consumed. For example, this happens through adaptive learning platforms that personalize content sequencing, through intelligent tutoring systems that offer real-time feedback, and through learning analytics that reshape curriculum design on the basis of student interaction patterns (Chen et al., 2022; Holmes et al., 2019). This transformation is not limited to STEM fields but has profoundly impacted language education, creating new opportunities for personalized and immersive learning (Kinsemolu & Onyeaka, 2025; Huang et al., 2023).

Bibliometric analyses indicate a growing interest in specific AI applications across various disciplines. These applications include intelligent tutoring systems, automated essay scoring, conversational agents (chatbots), predictive analytics for at-risk students, and natural language processing tools for content generation. Such interest ranges from engineering education (Liu et al., 2025) to library sciences (Borgohain et al., 2024). In the specific context of language learning, the focus has moved from simple translation tools to complex, adaptive systems that facilitate genuine communication (Rahman et al., 2024). This review examines the theoretical and empirical foundations of AI in education to contextualize the present study, which examines the success of AI-integrated English courses among adult learners over a four-month period.

To understand why AI is effective, one must look at the underlying learning theories. Traditional cognitivism and the principle of “animal–environment mutuality” (Costall, 2004) suggest that learning is deeply influenced by the interaction between the learner and their surroundings. In the digital age, this environment is becoming increasingly virtual. Maxwell (2006) argues for re situating constructionism in virtual environments, a concept that aligns with modern “Intelligent Educational Environments”, which model learner behavior to optimize outcomes (Terzieva et al., 2025).

This shift from physical to virtual and intelligent learning environments also touches on psychological and cultural dimensions. Guimaraes (2021) discusses semiotic-cultural constructivism, which helps explain how learners construct meaning in new environments. Furthermore, Khalili (2025) emphasizes that applying learning theories in interprofessional education can foster identity development, which is crucial for adult learners who often view language acquisition as a part of their professional growth. Sherman et al. (2025) further note that teachers act as the nexus between these theories and practical application, particularly for multilingual learners. The most prominent application of AI in the context of the current study is the use of chatbots and conversational agents. A meta-analysis by Wu and Yu (2023) confirmed that AI chatbots significantly improve student learning outcomes. These tools serve as conversation partners that are available 24/7, breaking the time constraints of traditional classrooms (Belda-Medina & Calvo-Ferrer, 2022). Research shows that these interactions are effective for various demographics. For instance, Haristiani et al. (2019) demonstrated the utility of “Gengobot” for

grammar, whereas Yang et al. (2022) reported that AI partners significantly benefit EFL speaking classes. Crucially, the nature of the interaction matters. Luke et al. (2017) compared human and chatbot partners and reported that while human interaction is valuable, chatbots can sustain interest through immediate responsiveness. This finding is supported by that of Lin and Mubarak (2021), who reported that combining mind mapping with chatbots enhanced flipped classroom experiences. However, the efficacy of these tools depends on how they are queried. Yang et al. (2025) reported that “asking questions matters,” with different effects when learners dialog with peers versus ChatGPT or intelligent tutoring systems. Similarly, Jeon (2022) explored these affordances specifically in EFL classrooms and reported that the perceived utility of the chatbot dictates the learner’s level of engagement.

A key advantage of AI, contributing to the “meaningful success” observed in the experimental group of this study, is personalization. Hardaker and Glenn (2025) and Chae et al. (2024) argue that AI platforms allow for a level of customization that human instructors cannot manage for large groups. By using learning analytics, educators can track pathways and outcomes with precision. This trend of personalization has been developing for twenty years (Chen et al., 2021). Modern systems, such as the visual programming environments explored by Chang et al. (2024), allow for adaptive models that respond to the learner’s cognitive state. In the context of English writing, Li et al. (2024) reported that natural language processing (NLP) algorithms can simulate e-learning environments that provide instant, corrective feedback, which is essential for adult learners seeking to correct fossilized errors. Recent research has also examined how learners adopt generative AI tools in educational contexts. Ittefaq et al. (2025) reported that international students’ perceived value and positive attitudes toward generative AI significantly influence their adoption of these tools for language learning. Similarly, Al Ghaithi, Behforouz, and Al-Hanishi (2026) reported that the use of Copilot AI in higher education settings improved grammar acquisition among English learners, particularly when students received immediate, context-sensitive feedback. These findings align with the focus of the present study on the relationships among learner perceptions, AI adoption, and language proficiency outcomes.

Adult learners often suffer from foreign language anxiety (FLA). Several studies included in this review suggest that AI is a solution to this emotional barrier. Bakkar (2019) investigated the concept that the home use of speech-enabled AI can mitigate anxiety, as the “judgment” of a machine is perceived as less threatening than that of a human. El Shazly (2021) confirmed this in a case study, showing that AI improved speaking performance by lowering anxiety levels. This reduction in anxiety is correlated with a greater willingness to communicate (WTC). Tai and Chen (2020) reported that tools such as Google Assistant encouraged adolescent learners to speak more, a finding transferable to adult contexts. Furthermore, positive psychology plays a role; Miao et al. (2025) reviewed empirical research and reported that positive emotional states facilitated by supportive tech environments enhance second language (SL) acquisition. However, Ho et al. (2022) and Kasap & Babayigit (2026) warn that we must also consider the attitudes of learners (specifically Gen Z and younger adults) toward nonconscious data collection, as “emotional AI” becomes more prevalent. AI often works in tandem with other emerging technologies. Divekar et al. (2022) and Godwin-Jones (2023) discuss the potential of “extended reality” (XR) and the metaverse, where AI bots populate immersive worlds to provide naturalistic language practice. These technologies are trending across education. Dogru et al. (2025) provided a bibliometric analysis of AR in STEM, and Rojas-Sanchez et al. (2022) did the same for virtual reality (VR), indicating a cross-disciplinary shift toward immersive learning that supports the findings of language-specific studies. The success of AI in language learning is mirrored in, and supported by, research in other fields. This cross-pollination of ideas validates the experimental approach.

- **Decision Making:** Lagzi et al. (2025) explored data-driven decision-making, which parallels how AI tutors decide which content to present to a learner next.
- **Assessment:** Hang et al. (2022) used analytical hierarchy processes to assess English teaching efficiency and demonstrated that AI can be used for evaluation as well as instruction.
- **Higher Education:** Klimova et al. (2023) and Yu and Nazir (2021) focus on university levels and note that 5G and AI transform situational teaching.
- **Music and Arts:** Wei et al. (2022) reported similar benefits in music education, suggesting that AI’s ability to teach patterns is universal across disciplines.
- **Social Media:** Lee (2023) analyzed Instagram and TikTok and reported that these algorithm-driven (AI) platforms offer unique “affordances” for language learning through short-form content.

The goal of AI is not to replace the teacher but to augment them. Ji et al. (2023) reviewed conversational AI with a focus on collaboration with human teachers, whereas Sankaran et al. (2022) proposed models for measuring the performance of human–AI collaborative processes. Wang et al. (2023) used cluster analysis to show that the quality of human–AI interactions is ultimately important for learning outcomes. However, this integration presents ethical challenges. Goldenthal et al. (2021) reported that “Not all AIs are equal,” and accessibility varies, potentially creating digital divides. Furthermore, as AI becomes ubiquitous, security becomes paramount. Al Siam et al. (2025) reviewed the impact of AI on cybersecurity, a relevant concern for educational platforms handling student data. Additionally, ethical considerations regarding information accuracy are raised by Oliveira et al. (2024) in the medical field and Sit et al. (2020) regarding student perceptions of AI, which underscores the need for reliable AI content in language courses.

Finally, in the context of inclusivity, Morris et al. (2021) and Chien et al. (2025) discuss how technology must be designed to avoid prejudice and assist distinct populations (such as the elderly or those with autism), ensuring that AI language tools are inclusive for all adult learners. The literature provides robust support for the integration of AI in education. From reducing anxiety (Bakkar, 2019) to providing adaptive feedback (Hwang & Zhang, 2024) and facilitating translation (Deng & Yu, 2022), the mechanisms underlying success are well documented. However, many reviews (Xie et al., 2019; Kundu & Bej, 2025) call for more empirical evidence on specific interventions.

This study used a strict experimental design with 72 adult learners over four months. The study was conducted in 2025. There were 36 learners in the experimental group and 36 in the control group. This design directly answers the need for more evidence in the field. The study follows the systematic reporting standards for reviews (Page et al., 2021). It also contributes useful data to the growing research encouraged by scholars such as Cai et al. (2025) on how AI can work across different subjects. The experimental group showed clear success. This success matches the broad agreement in the research literature that AI changes language learning from a passive activity into one that is dynamic, personal, and very effective.

Theoretical Framework

This study is guided by an integrated theoretical framework that combines three established theories to explain how AI affordances influence adult language learning outcomes. The framework addresses four key constructs: foreign language anxiety, learner autonomy, feedback processes, and language proficiency.

Krashen's Effective Filter Hypothesis

The first component is Krashen's affective filter hypothesis (1982), which proposes that anxiety acts as a mental barrier or "filter" that blocks language acquisition. When the filter is high (due to fear, embarrassment, or low confidence), comprehensible input cannot reach the language acquisition device of the brain. When the filter is low, acquisition proceeds naturally. AI tools, specifically voice assistants and chatbots, are hypothesized to lower the affective filter because they are perceived as non-judgmental. Unlike human teachers or peers, a machine does not evaluate, laugh, or lose patience. This perceived safety reduces foreign language anxiety, allowing more input to be processed. In the present study, compared with traditional instruction, the AI intervention is expected to significantly reduce anxiety levels.

Andragogy (Knowles, 1984)

The second component is Knowles' theory of andragogy, which describes how adults learn differently from children. Adults prefer self-directed learning (control over what and when they learn), problem-centered content (immediate relevance to real life), and intrinsic motivation (learning for personal growth rather than external rewards). AI tools align with all three principles. First, AI provides on-demand, 24/7 access, giving adult learners control over the pace and timing of their practice, thereby enhancing learner autonomy. Second, AI can generate personalized content on the basis of a learner's profession or interests (for example, a nurse reading medical dialogs), making learning problems-centered and immediately applicable. Third, the ability to see immediate progress through AI feedback sustains intrinsic motivation. In the present study, the AI intervention is expected to increase learner autonomy and engagement.

Feedback Loop Model (Li et al., 2024; Hattie & Timperley, 2007)

The third component is the feedback loop model, which emphasizes that the timing and specificity of feedback determine its effectiveness. Traditional classroom feedback is often delayed (days after submission), generally ("good job"), and single-loops (error correction without explanation). AI enables synchronous, specific, and iterative feedback. Learners receive immediate corrections the moment they make an error, and they can revise multiple times in one session (a tight feedback loop). This iterative process forces learners to notice the gap between their current interlanguage and the target language, leading to faster internalization of grammar rules and vocabulary. In the present study, the AI intervention is expected to produce significantly greater proficiency gains, particularly in writing and speaking, where immediate feedback is most critical.

Integration of the Three Theories

These three theories work together in a sequential chain. First, the non-judgmental nature of AI lowers the affective filter, reducing foreign language anxiety. Second, reduced anxiety and on-demand access increase learner autonomy, as adults feel safe and in control. Third, immediate, iterative feedback accelerates the development of language proficiency, especially in terms of productive skills. The qualitative strand of this study explores whether learners' reported experiences match this theoretical chain.

Methodology

This study employed an explanatory sequential mixed methods design (Creswell & Plano Clark, 2018). This two-phase approach was selected to first quantify the impact of the intervention and then use qualitative data to explain the mechanisms behind the quantitative outcomes.

Phase 1 (Quantitative): A true experimental pre-test post-test control group design with random assignment was implemented to measure the causal impact of AI integration on English proficiency and foreign language anxiety levels.

Phase 2 (Qualitative): Semistructured interviews were conducted with a purposive subsample of the experimental group to explore participants' subjective experiences and perceptions.

Participants and Setting

The participants were 72 adult learners enrolled in a general English as a foreign language program at a private language institute. All participants were screened to confirm a CEFR (Common European Framework of Reference for Languages) A2/B1 pre-intermediate proficiency level and were aged between 18 and 45 years.

After the pre-test, the participants were randomly assigned to one of two conditions using a computer-generated random number sequence created in Microsoft Excel. The assignment sequence was concealed until after the pre-test to prevent bias. The two conditions were as follows:

- Experimental Group (n = 36): Received AI-integrated blended instruction.
- Control Group (n = 36): Received traditional textbook-based instruction.

For the qualitative phase, a purposive sample of 10 participants was selected from the experimental group to capture diverse perspectives. Selection was based on post-test performance to include both high achievers (top 25 percent) and average improvers (middle 50 percent), ensuring a range of experiences with the AI intervention.

Table 1

Demographics of the Qualitative Interview Participants

Pseudonym	Age	Profession/Background	Ethnicity	Posttest Performance Group
P01	45	Accountant	Turkish	Average Improver
P02	41	Teacher	Kurdish	High Achiever
P03	35	Architect	Turkish	Average Improver
P04	34	Engineer	Turkish	High Achiever
P05	39	Sales Manager	Kurdish	Average Improver
P06	29	IT Specialist	Turkish	High Achiever
P07	22	University Student	Turkish	Average Improver
P08	30	Marketing	Kurdish	High Achiever
P09	28	Nurse	Turkish	Average Improver
P10	31	Project Coordinator	Turkish	High Achiever

Instruments and Materials

English Proficiency Test

The researcher developed a comprehensive English proficiency test aligned with the A2/B1 level of the Common European Framework of Reference for Languages. The test measured four language skills: listening (25 items), reading (25 items), speaking (25 points), and writing (25 points). The total possible score was 100.

In the listening section, the participants listened to short dialogs and answered multiple-choice questions about their main ideas and details. For the reading section, participants read two short passages (each approximately 200 words) and answer multiple-choice and true false questions. For the speaking section, participants responded to five prompts (for example, “Describe your daily routine” or “Order food at a restaurant”). The responses were audio recorded and later rated by two independent EFL instructors who were blinded to the group assignment. In the writing section, participants wrote a short paragraph (50 to 80 words) in response to a prompt such as “Write an email to a friend about your weekend.” The two same instructors rated the writing samples using a scoring rubric that included grammar (40 percent), vocabulary (30 percent), coherence (20 percent), and task completion (10 percent).

Interrater reliability was calculated using the intraclass correlation coefficient for absolute agreement. The coefficient was 0.89 for speaking and 0.91 for writing, indicating good to excellent reliability. Content validity was established by three senior EFL curriculum designers, who confirmed that the test items matched the A2/B1 level descriptors. The full test showed high internal consistency (Cronbach’s $\alpha = 0.85$).

Foreign Language Anxiety (FLA) Scale

Anxiety levels were measured using a 20-item scale adapted from the widely validated Foreign Language Classroom Anxiety Scale (Horwitz et al., 1986). The original scale contains 33 items. The adaptation process involved three steps.

First, redundant or context-specific items were removed. For example, items that referred exclusively to “in the classroom” were either dropped or rephrased to apply to general language learning situations because the AI intervention occurred partly outside the classroom. Second, the wording of some items was simplified to suit adult learners at the A2/B1 level. Third, the adapted scale was piloted with 15 adult learners who did not participate in the main study. On the basis of their feedback, three items were further revised for clarity.

The final 20-item scale preserved the three original dimensions of the FLCAS: communication apprehension (8 items), test anxiety (6 items), and fear of negative evaluation (6 items). A sample item for communication apprehension is “I feel very nervous when speaking English with a teacher.” A sample item for fear of negative evaluation is “I worry that other learners will laugh at me when I speak English.” A sample item for test anxiety is “I become anxious when I do not understand every word the teacher says.”

Responses were collected on a 5-point Likert scale where 1 indicated strongly disagree and 5 indicated strongly agree. The adapted scale showed good internal consistency in this study (Cronbach’s $\alpha = 0.88$). A confirmatory factor analysis was conducted to verify that the three-factor structure was preserved. The fit indices were acceptable: comparative fit index = 0.92 and root mean square error of approximation = 0.06. These values confirm that the shortened version retained the original theoretical dimensions.

3. Semi-Structured Interview Protocol

A protocol with open-ended questions was developed to explore learner perceptions. Sample questions included “Can you describe your experience practicing speaking with the AI assistant compared to a classmate?” and “How did receiving instant feedback on your writing affect your learning process?”

AI tools and protocols

I chose two AI tools for different reasons. For text-based interactions, I used ChatGPT (GPT-4o) because it is good at providing grammar feedback and rewriting sentences. It can also create reading passages that match the A2/B1 level of my learners. Many studies have shown that ChatGPT helps language learners improve their writing skills through immediate, personalized feedback. For voice-based practices, I selected Google Assistant. The main reasons were cost and accessibility. Google Assistant is free and works on most smartphones; thus, learners can practice at home without buying extra software. It also provides automatic speech recognition and basic pronunciation checks, which helped my participants build speaking confidence. I gave the experimental group a standardized guide to ensure that everyone used the tools in the same way. This guide included example prompts for conversation scenarios, specific commands for writing feedback (for example, “Check the grammar and suggest improvements for this sentence”) and instructions for generating personalized reading content.

Procedure

Phase 1: Quantitative Intervention (16 Weeks)

Both groups met for 4 hours of instructor-led instruction per week, covering identical core curriculum objectives. The control group used a standard textbook and engaged in pair or group work with peers and the instructor for practice and feedback. The experimental group followed a blended learning model in which the core instruction was augmented by structured AI integration. The intervention lasted 16 weeks and followed a weekly schedule, as shown in Table 2 (below).

Weekly AI Integration Protocol for the Experimental Group

Each week, participants in the experimental group completed three mandatory AI-based activities outside of class hours, totaling approximately 90 minutes per week.

1. **Voice Practice with a Google Assistant (20 minutes, twice per week):** Participants engaged in role play conversations using predefined scenarios (for example, ordering food at a restaurant, making a hotel reservation, or asking for directions). The exact prompts were provided in a standardized guide. Example prompt: *"Ask Google Assistant to pretend to be a waiter. You will order a meal. Say: 'I would like to order chicken and rice please.'"* Participants repeated each scenario at least three times, focusing on pronunciation and fluency. No teacher was present during these sessions, but the instructor reviewed anonymised interaction logs weekly.
2. **Writing Practice with ChatGPT (GPT-4o) (30 minutes, once per week):** Participants wrote three to five sentences on a given topic (matching the week's textbook theme). They then used the following standardized prompt for feedback: *"Check the grammar and suggest improvements for this sentence: [student's sentence]."* Participants were required to revise their sentences on the basis of the AI feedback and submit both the original and revised versions to the instructor. The instructor reviewed the revisions and addressed common errors in class.
3. **Personalized Reading Generation (20 minutes, once per week):** The participants asked ChatGPT to generate a short text at the A2/B1 level about their own profession or personal interest. The standardized prompt was *"Write a short text at A2/B1 English level about [topic]. Include 5 multiple-choice questions about the text."* Participants read the text, answered the questions, and then checked their answers with the AI.

Teacher Mediation and Monitoring

The instructor (the researcher) performed three monitoring activities each week. First, the instructor reviewed a random sample of 20 percent of all AI interaction logs (with participant consent) to check for appropriate use and technical problems. Second, the instructor dedicated 15 minutes of class time each week to discuss common errors or difficulties observed in the AI logs. Third, the instructor held a short weekly check-in with each participant during class to ask about their AI experience and address any frustration.

Intervention Fidelity

To ensure that the intervention was delivered as planned, two fidelity measures were used. First, each participant completed a simple checklist after every AI session, recording the date, the prompts used, and whether any technical issues occurred. Second, the researcher observed 10 percent of all AI sessions (randomly selected) either in person or through screen sharing to confirm that participants followed the standardized prompts. Adherence was high (94 percent of checklists confirmed full compliance).

Table 2: Weekly AI Integration Schedule for the Experimental Group (16 Weeks)

Week	Voice Practice (Google Assistant)	Writing Practice (ChatGPT)	Reading Practice (ChatGPT)	In Class Teacher Mediation
1 -2	Basic greetings and introduction	3 sentences about self	Text on learner's job	Review of common grammar errors
3 -4	Ordering food and drinks	5 sentences about daily routine	Text on learner's hobby	Pronunciation feedback
5 -6	Asking for directions	5 sentences describing a past event	Text on learner's city	Past tense correction
7 -8	Making a hotel reservation	5 sentences expressing opinion	Text on learner's favorite film	Opinion language review

Week	Voice Practice (Google Assistant)	Writing Practice (ChatGPT)	Reading Practice (ChatGPT)	In Class Teacher Mediation
9 -10	Shopping and returning items	5 sentences comparing two things	Text on learner's country	Comparative structures
11- 12	Doctor's appointment role play	5 sentences using future tense	Text on learner's future plans	Future tense practice
13 -14	Job interview role play	5 sentences describing a problem	Text on learner's professional field	Problem solution language
15 -16	Review of all scenarios	Review and free writing	Learner chooses any topic	Overall pro

4.5 Data Analysis

Quantitative Analysis: The data were analyzed using SPSS v.28. Independent-samples t-tests were used to compare post-test proficiency and FLA scores between groups. Paired-samples t-tests were used to assess within-group changes from pre-- to post-test. To more precisely evaluate the impact on anxiety while controlling for baseline levels, an analysis of covariance (ANCOVA) was conducted with post-intervention FLA score as the dependent variable and pre-intervention FLA score as the covariate.

Qualitative Analysis: The interview transcripts were analyzed using reflexive thematic analysis (Braun & Clarke, 2006, 2019). The process involved (1) familiarization with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report. Coding was conducted by two researchers to enhance trustworthiness, with discrepancies resolved through discussion. To improve the validity of the qualitative findings, two additional strategies were employed. First, member checking was performed by returning a summary of the identified themes to five participants, who confirmed that the interpretations matched their experiences. Second, a reflexive journal was maintained throughout the analysis to document the researcher's assumptions and decisions, thereby reducing potential bias.

Exclusion criteria

An attendance threshold of 80% was enforced. All 72 initial participants met this criterion, and no data were excluded.

Ethical Considerations

Informed consent was obtained from all participants prior to the start of the study. Participants were informed that the research aimed to compare different teaching methods but were not informed of the specific hypotheses or that one group was designated a control group. This was done to minimize demand characteristics. All the participants were assured that their participation was voluntary and that they could withdraw at any time without penalty. No participants withdrew during the 16-week intervention.

All the data were anonymized. Each participant was assigned a code number, and personal identifiers were removed from all datasets. The interview audio recordings were stored on a password-protected computer and were destroyed immediately after transcription. Written consent forms were kept in a locked filing cabinet accessible only to the researcher.

Findings

This section presents the results of the data analysis in two parts. First, the quantitative findings regarding English proficiency and foreign language anxiety (FLA) are presented using descriptive and inferential statistics. Second, the qualitative findings from the semistructured interviews are analyzed to explain the participants' perceptions of the AI-integrated learning process.

Quantitative Findings

This section details the statistical analysis of English proficiency and anxiety data collected from 72 participants (N = 72). Analyses were conducted using SPSS version 28 and included independent samples t- tests, paired samples t- tests, and two one-way ANCOVAs.

To ensure the validity of the experiment, it was crucial to confirm that both groups started with equivalent levels of English proficiency and foreign language anxiety. Independent samples t-tests conducted on the pre-test scores revealed no statistically significant differences between the control and experimental groups for overall proficiency or for initial anxiety levels, confirming group homogeneity at the outset of the study (see Table 2).

Table 3: Comparison of Pre-Test Total Proficiency and Anxiety Scores

Group	N	Proficiency Mean (M)	Proficiency SD	Anxiety Mean (M)	Anxiety SD
Control	36	48.20	5.45	4.10	0.62
Experimental	36	48.65	5.80	4.05	0.58

*Note: Proficiency max score = 100. Anxiety was measured on a 1-5 scale (1 = low anxiety, 5 = high anxiety). Independent sample t-tests revealed no significant differences: proficiency $t(70) = -0.34$, $p = .732$; anxiety $t(70) = 0.35$, $p = .728$. *

After the 16-week intervention, an independent samples t- test was conducted to compare the final learning outcomes. As detailed in Table 3, compared with the control group ($M = 61.15$, $SD = 6.20$), the experimental group achieved a significantly higher mean proficiency score ($M = 76.40$, $SD = 6.85$). The difference of 15.25 points was statistically significant, indicating that the AI-integrated method was considerably more effective.

Table 4: Comparison of Post-Test Total Proficiency Scores with Effect Size and Confidence Interval

Group		Mean (M)	D	p	Cohen's d	95% CI for Mean Difference
Control	6	61.15	.20	.84	<.001	2.32 [12.89, 17.61]
Experimental	6	76.40	.85			

To control for pre-test proficiency differences, a one-way ANCOVA was conducted. After adjusting for pre-test scores, the experimental group (adjusted mean = 76.12, $SE = 1.02$) scored significantly higher than the control group (adjusted mean = 61.48, $SE = 1.02$), $F(1, 69) = 98.45$, $p < .001$, partial $\eta^2 = 0.59$. This represents a large effect size. The 95% confidence interval for the adjusted mean difference was [12.89, 17.61]. To identify where the AI tools had the greatest effect, post-test scores were analyzed by sub-skill. Independent-samples t-tests confirmed that the experimental group significantly outperformed the control group in all four domains (see Table 5). The magnitude of the difference varied, with the greatest effects observed for productive skills.

Table 5: Post-Test Scores by Skill Domain with Between-Group Comparisons, Effect Sizes, and Confidence Intervals

Speaking			
Group	M (SD)	t	p
Control	14.50 (2.10)	10.82	<.001
Experimental	19.80 (1.95)		
Diff [95% CI]	+5.30 [4.32, 6.28]		
Cohen's d	2.55		

Writing			
Group	M (SD)	t	p
Control	15.10 (2.30)	7.69	<.001
Experimental	19.20 (2.15)		
Diff [95% CI]	+4.10 [3.01, 5.19]		
Cohen's d	1.81		
Reading			
Group	M (SD)	t	p
Control	16.20 (2.50)	5.25	<.001
Experimental	19.20 (2.40)		
Diff [95% CI]	+3.00 [1.86, 4.14]		
Cohen's d	1.23		
Listening			
Group	M (SD)	t	p
Control	15.35 (2.20)	5.55	<.001
Experimental	18.20 (2.10)		
Diff [95% CI]	+2.85 [1.82, 3.88]		
Cohen's d	1.30		

Analysis of Skill-Based Findings:

1. **Speaking (Greatest Improvement):** The greatest performance gap was observed in Speaking skills (mean difference = +5.30, $p < .001$, Cohen's $d = 2.55$). This substantial effect suggests that the use of voice-enabled AI agents provided the high-frequency, low-anxiety oral practice that is often logistically constrained in traditional classrooms.
2. **Writing (High Improvement):** The second-largest gain was in Writing (mean difference = +4.10, $p < .001$, Cohen's $d = 1.81$). This correlates directly with the methodology of iterative AI feedback, which allowed students to engage in immediate drafting and correction cycles.
3. **Reading and Listening (Significant Improvement):** While the AI group performed significantly better in receptive skills (Reading: +3.00, $p < .001$; Cohen's $d = 1.23$; Listening: +2.85, $p < .001$; Cohen's $d = 1.30$), the effect sizes were somewhat smaller than those for productive skills. This finding indicates that while personalized AI content generation is effective, traditional methods also support the development of receptive competencies.

The primary objective was to measure the intervention's impact on learners' psychological experience. Paired samples t-tests within each group revealed a significant reduction in anxiety for the experimental group. For the Experimental Group, the reduction from the pre-test ($M = 4.05$, $SD = 0.58$) to the post-test ($M = 2.18$, $SD = 0.62$) was large and significant, $t(35) = 12.47$, $p < .001$, Cohen's $d = 3.12$, 95% CI for the mean change $[1.56, 2.18]$. For the control group, the reduction from the pre-test ($M = 4.10$, $SD = 0.62$) to the post-test ($M = 3.85$, $SD = 0.60$) was not significant, $t(35) = 1.89$, $p = .067$.

To isolate the effect of the intervention on anxiety while controlling for initial levels, ANCOVA was conducted with post-test anxiety as the dependent variable, group anxiety as the independent variable, and pre-test anxiety as the covariate. The assumptions for this ANCOVA were checked and satisfied: normality (Shapiro–Wilk $p > .05$), homogeneity of variances (Levene's $p > .05$), and homogeneity of regression slopes (interaction $p > .05$).

Table 6: Foreign Language Anxiety (FLA) Scores Pre- and Post-Intervention with ANCOVA Results and Confidence Intervals

Group	Pre-Test Anxiety (M)	Post-Test Anxiety (M)	Within-Group Change	ANCOVA Results (Post-Test, controlling for Pre-Test)
Control	4.10	3.85	-0.25	$F(1, 69) = 42.37$, $p < .001$, $\eta^2 = 0.38$, 95% CI for adjusted mean difference $[1.12, 1.98]$
Experimental	4.05	2.18	1.87	

*Note: Anxiety measured on a 1-5 scale (1 = low anxiety, 5 = high anxiety). *

The analysis revealed a statistically significant effect of group membership on post-test anxiety after accounting for pre-test differences, $F(1, 69) = 42.37$, $p < .001$, partial $\eta^2 = 0.38$, indicating a large effect size. The 95% confidence interval for the adjusted mean difference between groups was $[1.12, 1.98]$. This significant reduction in anxiety for the Experimental Group corresponds with their higher proficiency outcomes, supporting the hypothesis that AI integration addresses both cognitive and affective barriers to adult language acquisition.

5.2 Qualitative findings

To complement and explain the quantitative results, semistructured interviews were conducted with a purposive sample of 10 participants from the experimental group (the demographic data are presented in Table 5). Thematic analysis of the interview transcripts revealed four interconnected themes that elucidate *how* the AI-integrated method influenced the learning process: (1) the creation of a nonjudgmental safe space, (2) the critical role of immediate feedback, (3) empowerment through learner autonomy, (4) sustained engagement via personalization and (5) negative cases.

Theme 1: AI as a Nonjudgmental Safe Space

The most prominent theme concerned the psychological safety afforded by AI interaction. The participants consistently described the AI as a patient and pressure-free practice partner, sharply contrasting it with the anxiety of speaking before peers or instructors. This perceived absence of social judgment was pivotal in reducing foreign language anxiety and encouraging oral experimentation. P04 noted, *"In class, I am always afraid my grammar is bad, so I stay quiet. However, the AI does not judge me... It made me brave enough to speak in class later."* Similarly, P09 valued the privacy, stating, *"I could practice my pronunciation in my living room without worrying that I sounded foolish."* This sentiment frames the AI as a low-stakes rehearsal space, effectively lowering the affective filter and building confidence for real-world communication.

Theme 2: The Transformative Value of Immediate Feedback

The participants emphasized that the timing and nature of AI feedback fundamentally altered their learning cycle. Unlike the delayed feedback typical in traditional settings, the AI's instant corrections created a tight, iterative loop that facilitated immediate error recognition and correction. This was particularly valued in writing. P02 explained, *"Usually, I get my homework back two days later, and I forget why I made the mistake. With the AI... I fix the error the second I make it. It explains 'why' I was wrong immediately."* P07 added that this consistency led to breaking persistent error patterns: *"I stopped making the same tense mistakes because the AI caught them every single time."* This finding directly supports the quantitative finding of significant gains in writing accuracy.

Theme 3: Enhanced Learner Autonomy and Accessibility

For adult learners managing professional and personal commitments, the asynchronous, on-demand support of AI tools was a key factor in sustained engagement. This theme highlights a shift in agency, where learners felt in control of the pace and

timing of their study. P05, whose irregular work schedule posed a challenge, stated, *"The chatbot is there at midnight... It made me feel in control of my learning."* P08 emphasized how this accessibility prevented discouragement: *"I could clear up my confusion immediately, so I did not fall behind. It felt like the class was moving at my speed."* The AI thus functioned as a perpetual tutor, supporting self-regulated learning outside the classroom.

Theme 4: Engagement Fueled by Personalization

The final theme centered on the ability to tailor learning content to personal and professional interests. Moving beyond generic textbook material, the learners used AI to generate relevant texts and scenarios, which increased their intrinsic motivation and perceived relevance. P06, an IT specialist, found standard topics disengaging: *"I used the AI to generate English articles about coding and technology at my level. Because I was interested in the topic, I actually wanted to read more."* P03, an architect, directly applied the tool to her career needs: *"I asked it to create dialogs between an architect and a client. It was exactly the English I needed for my job."* This personalization aligns with andragogical principles, making learning directly applicable and thereby sustaining effort and interest.

In summary, the qualitative data provide a coherent explanatory model for quantitative success. The AI tools created a psychologically safe environment that reduced anxiety, enabled a powerful immediate feedback loop that accelerated skill acquisition, offered unprecedented accessibility that fostered autonomy, and delivered personalized content that maintained engagement. Together, these four factors describe the experiential mechanisms behind the "meaningful success" observed in the statistical data.

Theme 5: Negative Cases

Although the majority of participants reported positive experiences with the AI tools, three learners (P03, P07, and P09) described challenges or frustrations that qualify as negative or discrepant cases. Their experiences are reported here to provide a balanced interpretation.

P03, an architect and average improver, stated that the voice recognition feature of Google Assistant sometimes misunderstood her accent, especially when she spoke words with Turkish pronunciation patterns. She said, "I would say a word correctly three times, but the Assistant kept hearing something else. That made me feel like my pronunciation was worse than I thought." This frustration led her to reduce her voice practice in weeks 6 to 8, although she resumed later with more patience.

P07, a university student and average improver, reported that the AI occasionally gave inconsistent feedback on the same grammar error. He said, "One time ChatGPT corrected my use of 'in' to 'at'. The next day, it left the same sentence unchanged and said it was fine. I was confused about which rule to follow." This inconsistency caused him to doubt the reliability of the AI, and he expressed a preference for asking the human teacher for confirmation.

P09, a nurse and average improver, described technical difficulties related to internet access. She explained, "My home internet connection is not always stable. Sometimes the AI would stop responding in the middle of a conversation, and I lost my work. That was very discouraging." Despite these issues, P09 completed all the required activities by using the institute's computers after class hours.

No participant expressed a desire to return fully to traditional methods. However, these discrepant cases highlight three important limitations of AI integration: variable speech recognition for nonnative accents, occasional inconsistency in feedback, and a reliance on stable internet connectivity. These findings suggest that AI tools work best when supplemented with human teacher support, offline backup options, and clear guidance on how to handle technical failures or conflicting feedback.

6. Discussion

The primary objective of this study was to investigate the effectiveness of AI-integrated instruction on the English language proficiency and foreign language anxiety of adult EFL learners. By employing a mixed methods design with 72 participants, this research sought to determine whether replacing traditional practice methods with generative AI tools could produce meaningful success in a private language institute setting. The findings of this study are promising but not definitive. The quantitative data revealed that the experimental group ($n = 36$) outperformed the control group ($n = 36$) in all four language domains: listening, speaking, reading, and writing. Furthermore, the experimental group exhibited a large and statistically significant reduction in anxiety levels. The qualitative data, derived from semistructured interviews, provided contextual explanations for these results, highlighting the themes of psychological safety, immediate feedback, and enhanced learner autonomy. However, because the study was conducted at a single institution in Turkey with a relatively small sample, these findings should be interpreted with caution. Replication studies with larger and more diverse populations are needed before strong claims about generalizability can be made.

A notable finding of this study was the difference in posttest proficiency scores between groups. The AI-integrated group achieved a mean score of 76.40, whereas it was 61.15 for the control group. While both groups improved, the learning curve of the AI group was steeper. This pattern is consistent with recent meta-analyses by Chiu et al. (2023) and Zhai et al. (2021), which suggest that AI tools can serve as effective knowledge scaffolds in language education. However, it is possible

that some of the observed gains were due to the novelty of the technology rather than the pedagogical quality of the AI itself. Longitudinal research is needed to determine whether these effects persist beyond the first few months of use.

The detailed breakdown of scores revealed that the greatest gains were in the productive skills of speaking and writing. This finding is meaningful because these skills are often the hardest to develop in traditional classrooms because of time limitations and large class sizes. The success of the experimental group may be attributed to the volume of deliberate practices made possible by AI voice agents. Bin Hady et al. (2023) argued that a primary barrier to adult fluency is the output gap: Learners understand more than they can say because they rarely obtain enough speaking time in a class of 20 or more students. In this study, AI voice agents acted as conversation partners available outside class hours, allowing each participant to practice individually. This high-frequency repetition may have helped learners move from controlled processing (thinking about each word) toward more automatic processing (fluent speech). However, it is also possible that learners simply enjoyed using the technology and therefore practiced more. The study design did not separate the effect of increased practice time from the effect of AI-specific features.

In writing, the success of the AI group supports the feedback loop theory proposed by Li et al. (2024). In the control group, feedback was delayed: students wrote a draft, submitted it, and received corrections days later. By that time, the opportunity for immediate learning had often passed. In contrast, the experimental group received synchronous feedback from the AI. As noted in the qualitative interviews, participants could identify and correct errors at the moment they occurred. This iterative drafting process may have helped learners notice the gap between their current interlanguage and the target language, leading to faster internalization of grammar rules (Hwang & Chang, 2023). Nonetheless, the study did not measure whether these gains were retained after the intervention ended. Future research should include delayed post tests to assess long-term retention.

The AI group also outperformed the control group in reading and listening, although the effect sizes were smaller than those for speaking and writing. While the control group relied on a standard textbook, the experimental group used AI to generate personalized reading materials. This finding is consistent with the principles of andragogy (Knowles, 1984), which suggest that adults learn best when content is directly relevant to their lives or careers. By prompting the AI to generate texts about their own professions (for example, engineering, nursing, or IT), participants engaged in what Fitria (2023) described as contextualized learning. When a learner reads a text about his or her own job, the cognitive load of understanding the context is reduced, allowing more attention to be focused on language features. However, the smaller improvement in receptive skills than in productive skills suggests that traditional instruction may also be reasonably effective for listening and reading. AI may offer added value primarily for skills that require active production and immediate feedback.

One of the more notable contributions of this study is the empirical evidence linking AI usage to a reduction in foreign language anxiety. The anxiety scores for the experimental group decreased from a pretest mean of 4.05 (high anxiety) to a posttest mean of 2.18 (low anxiety), a change that was statistically significant with a large effect size (partial $\eta^2 = 0.38$). The control group showed only a small, nonsignificant reduction from 4.10 to 3.85. These results suggest that AI integration may help lower the affective filter described by Krashen (1982).

The qualitative data strongly suggested that the primary driver of this reduction in anxiety was the nonjudgmental nature of the AI. Adult learners often fear losing face, that is, the embarrassment of making mistakes in front of peers or instructors. The participants' comments indicated that the AI was perceived as a rehearsal space where errors carried no social cost. These findings support the safe space hypothesis proposed by Hong (2023): the AI acts as a transitional object that is interactive like a human but lacks the capacity to judge, mock, or lose patience. When the fear of negative evaluation is reduced, learners may become willing to take more risks, try more complex sentences, and speak more frequently. However, it is also possible that the reduction in anxiety was partly due to the structured nature of the AI activities rather than the AI itself. A comparison with a structured self-study workbook would help isolate the effect of the AI's nonjudgmental character.

The negative correlation between anxiety and proficiency scores in this study is consistent with the view that anxiety acts as a cognitive block. The control group, which remained at higher anxiety levels, may have spent a portion of their cognitive resources managing stress rather than processing language. The experimental group, having offloaded that stress to the nonjudgmental AI, could dedicate more cognitive capacity to language acquisition. This interpretation aligns with Wei's (2023) assertion that the emotional benefits of AI are as important as pedagogical benefits. Nevertheless, the correlational nature of these findings does not establish causation. It remains possible that learners who were already more proficient experienced less anxiety, rather than the AI reducing anxiety directly.

6.1 Pedagogical Implications: A New Role for Teachers

The results of this study suggest a possible shift in the pedagogical model for adult EFL education. The success of the blended model implies that some tasks currently performed by teachers might be effectively supported by AI. However, the findings do not support the replacement of human teachers. Instead, they point toward a reconfiguration of roles. First, teachers may consider reducing the amount of class time spent on rote grammar drills or basic vocabulary checks, as AI appears capable of handling these tasks efficiently. This would free up class time for higher-level communicative tasks, cultural discussions, and activities that require human nuance, such as debate and negotiation. Second, the role of teachers may expand to include mentoring students on how to use AI tools effectively, including prompt engineering and critical

evaluation of AI outputs. Third, curriculum designers may consider integrating structured AI practice slots into lesson plans, recognizing that short periods of AI conversation can supplement limited classroom speaking time. These implications are tentative. The study did not measure teacher workload or the time required to monitor AI sessions. Institutions considering AI integration should be aware that initial setup, training, and monitoring require investment. Furthermore, not all teachers may feel comfortable with AI tools, and professional development is necessary.

6.2 Limitations and Future Research

While this study provides promising evidence for the efficacy of AI integration, several limitations must be acknowledged. First, the study was conducted over 16 weeks. This is sufficient to measure immediate gains but does not account for the novelty effect. Participant engagement may have been artificially high because the technology was new and interesting. Longitudinal studies spanning a full academic year are needed to determine whether motivation and learning effects persist once novelty fades.

Second, the sample consisted of 72 adult learners at a single private language institute in Turkey, all of whom had A2 or B1 proficiency levels and were aged 18 to 45. The results may not be generalizable to older adults, to learners in public education settings, or to learners from different cultural or linguistic backgrounds. Replication studies with larger, more diverse samples are essential.

Third, the study relied on specific generative AI models (ChatGPT GPT-^o and Google Assistant). Given the rapid pace of technological change, these specific tools may soon be outdated. However, the underlying principles (immediate feedback, personalization, and low stakes practice) are likely to remain relevant regardless of the software used. Future research should examine whether similar effects are observed with newer or different AI tools.

Fourth, there is the issue of AI hallucinations, that is, the generation of incorrect or nonsensical information. While this study focused on language form (grammar and syntax) rather than factual content, there remains a risk that AI could reinforce incorrect language patterns if not monitored. This study included teacher reviews of AI logs, but the frequency of AI errors was not systematically recorded. Future research should investigate the accuracy rates of AI tutors in complex linguistic scenarios and develop error correction protocols.

Fifth, the study assumed a baseline level of digital literacy. The participants were aged 18 to 45 years and were generally comfortable with smartphones and basic software. The results may not be generalizable to older adult learners or to individuals from lower socioeconomic backgrounds with limited access to technology (the digital divide). Future research should examine AI-assisted language learning with older adults and with populations who have limited prior exposure to digital tools.

Finally, the qualitative analysis, while following the reflexive thematic analysis, was conducted by two researchers from the same institution. Member checking was performed, but external audits or peer debriefing were not used. The inclusion of discrepant cases (reported in the findings) strengthens credibility, but the possibility of researcher bias cannot be fully ruled out.

7. Conclusion

This study provides promising evidence that the integration of artificial intelligence into EFL instruction can support improvements in both English proficiency and anxiety reduction for adult learners. Within the specific context of this 16-week intervention with 72 participants at a private language institute in Turkey, the AI-integrated approach addressed two common challenges in adult language learning: limited opportunities for practice and the fear of making mistakes. The qualitative insights suggest that AI may serve not only as a tool for efficiency but also as a tool for empowerment, giving learners greater control over the pace, content, and timing of their learning. However, these findings are based on a single institution with a relatively small sample. Replication studies with larger, more diverse populations and longer follow-up periods are needed before strong claims about generalizability can be made. The findings offer several practical recommendations for institutions considering AI integration. First, teacher training in prompt engineering and AI monitoring is essential. Second, institutions should develop ethical guidelines covering data privacy, acceptable use, and the disclosure of AI-generated content. Third, learner AI literacy should be taught explicitly through orientation modules. Fourth, privacy protocols must ensure compliance with data protection laws and offer data deletion options. Fifth, equitable access requires providing loaner devices, computer lab access, or offline alternatives for learners without reliable internet. Finally, institutions are encouraged to start with small-scale pilots before full implementation.

These recommendations are tentative and based on a single study, but they offer a starting point for responsible and effective AI integration. For educators and institutions, the results of this study suggest that AI integration is a promising direction worthy of further investigation, not a settled solution. Practical next steps include piloting AI tools in one or two classes, providing teacher training, developing clear guidelines for appropriate use, and monitoring outcomes carefully. Future research should examine long-term retention, the novelty effect, the effectiveness of AI with different age groups and proficiency levels, and the cost effectiveness of AI integration compared to other interventions, such as reduced class size or additional human tutoring. As AI technology continues to mature, its role in education may expand. The question for is now

not whether AI will be used but how it can be integrated in ways that are ethical, equitable, and pedagogically sound. This study offers one example of how that might be done, but many questions remain open for future inquiry.

Disclosure statement

No potential conflicts of interest were reported by the author(s). The author(s) declare that they have no relevant financial or nonfinancial interests to disclose. The author(s) utilized Gemini (a large language model by Google) solely for language editing and proofreading assistance in the preparation of this manuscript; the author(s) take full responsibility for the content.

References

- Akinsemolu A. A., & Onyeaka H. (2025). The Role of Artificial Intelligence in Transforming Language Learning: Opportunities and Ethical Considerations. *Journal of Language and Education*, 11(1), 148-152. <https://doi.org/10.17323/jle.2025.2211>
- Al Ghaithi, A., Behforouz, B., & Al-Hanshi, S. (2026). Empowering grammar acquisition through Copilot AI in higher education. *Journal of International Students*, 16(11), 339-356. <https://doi.org/10.32674/za5eta42>
- Al Siam, A., Alazab, M., Awajan, A., & Faruqui, N. (2025). A comprehensive review of AI's current impact and future prospects in cybersecurity. *IEEE Access*, 13, 14029-14050. <https://doi.org/10.1109/ACCESS.2025.3528114>
- Bakkar, B. B. (2019). The Writing Center's role in the academic life of an English foreign language student from an instructor's perspective. *Arab World English Journal*, 10(3), 32-42. <https://dx.doi.org/10.24093/awej/vol10no3.3>
- Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, 12(17), 8427. <https://doi.org/10.3390/app12178427>
- Bin-Hady, W. R. A., Al-Kadi, A., Hazaea, A., & Ali, J. K. M. (2023). Exploring the dimensions of ChatGPT in English language learning: A global perspective. *Library Hi Tech*, 41(5), 1303-1323. <https://doi.org/10.1108/LHT-05-2023-0200>
- Borghain, D. J., Bhardwaj, R. K., & Verma, M. K. (2024). Mapping the literature on the application of artificial intelligence in libraries (AAIL): A scientometric analysis. *Library Hi Tech*, 42(1), 149-179. <https://doi.org/10.1108/LHT-07-2022-0331>
- 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, C., Zhu, G., & Ma, M. (2024). A systemic review of AI for interdisciplinary learning: Application contexts, roles, and influences. *Education and Information Technologies*, 30(7), 9641-9687. <https://doi.org/10.1007/s10639-024-13193-x>
- Chae, J., Kim, M.-Y., Ryu, K., You, Y.-M., & Shin, Y.-H. (2024). Analysis of an AI-based chatbot platform for personalized learning. *Journal of Digital Contents Society*, 25(4), 1053-1068. <https://doi.org/10.9728/dcs.2024.25.4.1053>
- Chang, J. H., Wang, C. J., Zhong, H. X., Weng, H.-C., Zhou, Y.-K., Ong, H.-Y., & Lai, C.-F. (2024). Artificial intelligence learning platform in a visual programming environment: Exploring an artificial intelligence learning model. *Educational Technology Research and Development*, 72(2), 997-1024. <https://doi.org/10.1007/s11423-023-10323-z>
- Chen, X., Zou, D., Xie, H., & Cheng, G. (2021). Twenty years of personalized language learning: Topic modeling and knowledge mapping. *Journal of Educational Technology & Society*, 24(1), 205-222.
- Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two decades of artificial intelligence in education: Contributors, collaborations, research topics, challenges, and future directions. *Journal of Educational Technology & Society*, 25(1), 28-47.
- Chien, S. C., Yen, C. M., Chang, Y. H., Chen, Y. E., Liu, C. C., Hsiao, Y. P., Yang, P. Y., Lin, H. M., Yang, T. E., Lu, X. H., Wu, I. C., Hsu, C. C., Chiou, H. Y., & Chung, R. H. (2025). Use of artificial intelligence, internet of things, and edge intelligence in long-term care for older people: Comprehensive analysis through bibliometric, Google Trends, and content analysis. *Journal of Medical Internet Research*, 27, e56692. <https://doi.org/10.2196/56692>
- Chiu, T. K. F., Moorhouse, B. L., Chai, C. S., & Ismailov, M. (2023). Teacher support and student motivation to learn with artificial intelligence (AI) based chatbot. *Interactive Learning Environments*. Advance online publication. <https://doi.org/10.1080/10494820.2023.2172044>
- Costall, A. (2004). From Darwin to Watson (and cognitivism) and back again: The principle of animal-environment mutuality. *Behavior and Philosophy*, 32(1), 179-195.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Deng, X., & Yu, Z. (2022). A systematic review of machine-translation-assisted language learning for sustainable education. *Sustainability*, 14(13), 7598. <https://doi.org/10.3390/su14137598>

- Divekar, R. R., Drozdal, J., Chabot, S., Zhou, Y. L., Su, H., Chen, Y., Zhu, H. M., Hendler, J. A., & Braasch, J. (2022). Foreign language acquisition via artificial intelligence and extended reality: Design and evaluation. *Computer Assisted Language Learning*, 35(9), 2332-2360. <https://doi.org/10.1080/09588221.2021.1879162>
- Dogru, M. S., Yüzbaşıoğlu, F., & Arpacı, I. (2025). A bibliometric trend analysis of augmented reality in STEM education. *Innovations in Education and Teaching International*, 1-16. <https://doi.org/10.1080/14703297.2025.2483387>
- El Shazly, R. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, 38(3), e12667. <https://doi.org/10.1111/exsy.12667>
- Fitria, T. N. (2023). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. *ELT Forum: Journal of English Language Teaching*, *12*(1), 44-58. <https://doi.org/10.15294/elt.v12i1.64069>
- Godwin-Jones, R. (2023). Emerging spaces for language learning: AI bots, ambient intelligence, and the metaverse. *Language Learning & Technology*, 27(2), 6-27. <http://hdl.handle.net/10125/73501>
- Goldenthal, E., Park, J., Liu, S. X., Mieczkowski, H., & Hancock, J. T. (2021). Not all AI are equal: Exploring the accessibility of AI-mediated communication technology. *Computers in Human Behavior*, 125, 106975. <https://doi.org/10.1016/j.chb.2021.106975>
- Guimaraes, D. S. (2021). Where is semiotic-cultural constructivism in psychology heading? Contemporary reflections on the trajectory of an approach to cultural psychology. *Culture and Psychology*, 27(4), 523-538. <https://doi.org/10.1177/1354067X211017300>
- Hang, Y., Khan, S., Alharbi, A., & Nazir, S. (2022). Assessing English teaching linguistic and artificial intelligence for efficient learning using analytical hierarchy process and Technique for Order of Preference by Similarity to Ideal Solution. *Journal of Software*, 17(1-2), 31-39. <https://doi.org/10.1002/smr.2462>
- Hardaker, G., & Glenn, L. E. (2025). Artificial intelligence for personalized learning: A systematic literature review. *International Journal of Information and Learning Technology*, 42(1), 1-14. <https://doi.org/10.1108/IJILT-07-2024-0160>
- Haristiani, N., Danuwijaya, A. A., Rifa'i, M. M., & Sarila, H. (2019). Gengobot: A chatbot-based grammar application on mobile instant messaging as language learning medium. *Journal of Engineering Science and Technology*, 14(6), 3158-3173.
- Ho, M., Mantello, P., Ghotbi, N., Nguyen, M., Nguyen, H. T., & Vuong, Q. (2022). Rethinking technological acceptance in the age of emotional AI: Surveying Gen Z (Zoomer) attitudes toward nonconscious data collection. *Technology in Society*, 70, 102011. <https://doi.org/10.1016/j.techsoc.2022.102011>
- Hong, W. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation*, 5(1), 37-55. <https://doi.org/10.61414/jeti.v5i1.103>
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125-132. <https://doi.org/10.1111/j.1540-4781.1986.tb05256.x>
- Huang, X. Y., Zou, D., Cheng, G., Chen, X. L., & Xie, H. R. (2023). Trends, research issues and applications of artificial intelligence in language education. *Journal of Educational Technology & Society*, *26*(1), 112-131.
- Hwang, G. J., & Chang, C. Y. (2023). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, 31(7), 4099-4112. <https://doi.org/10.1080/10494820.2021.1952615>
- Hwang, G.-J., & Zhang, D. (2024). Effects of an adaptive computer agent-based digital game on EFL students' English learning outcomes. *Educational Technology Research and Development*. Advance online publication. <https://doi.org/10.1007/s11423-024-10396-4>
- Jeon, J. (2022). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*. Advance online publication. <https://doi.org/10.1080/09588221.2021.2021241>
- Ji, H., Han, I., & Ko, Y. (2023). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education*, 55(1), 48-63. <https://doi.org/10.1080/15391523.2022.2142873>
- Jowsey, T., Braun, V., Clarke, V., Lupton, D., Fine, M., & 414 Signatories. (2025). We reject the use of generative artificial intelligence for reflexive qualitative research. *Qualitative Inquiry*. Advance online publication. <https://doi.org/10.1177/10778004251401851>
- Ittefaq, M., Zain, A., Arif, R., Ahmad, T., Khan, L., & Seo, H. (2025). Factors influencing international students' adoption of generative artificial intelligence: The mediating role of perceived values and attitudes. *Journal of International Students*, 15(7), 127-154. <https://orcid.org/0000-0002-9653-508X>
- Khalili, H. (2025). Learning theories and their applications in interprofessional education (IPE) to foster dual identity development. *Journal of Interprofessional Care*, 39(3), 1-10. <https://doi.org/10.1080/13561820.2025.2496325>

- Kasap, S., & Babayiğit, M. V. (2026). Development and validation of the AI Addiction Scale (AIAS-10): Measuring compulsive and emotional dependence on AI tools among EFL learners. *Register Journal*, 19(1), 1–27. <https://doi.org/10.18326/register.v19i1.1-27>
- Klimova, B., Pikhart, M., Polakova, P., Cerna, M., Yayilgan, S. Y., & Shaikh, S. (2023). A systematic review on the use of emerging technologies in teaching English as an applied language at the university level. *Systems*, 11(1), 42. <https://doi.org/10.3390/systems11010042>
- Kohnke, L. (2023). L2 learners' perceptions of a chatbot as a potential partner for practicing L2 communication. *International Journal of Mobile Learning and Organization*, 35(1), 105-119. <https://doi.org/10.1504/IJMLO.2023.10053355>
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Kundu, A., & Bej, T. (2025). Transforming EFL teaching with AI: A systematic review of empirical studies. *International Journal of Artificial Intelligence in Education*. Advance online publication. <https://doi.org/10.1007/s40593-025-00470-0>
- Lagzi, M. D., Farkhondeh, F., Amoozad Mahdiraji, H., & Sakka, G. (2025). Exploring data-driven decision-making practices: A comprehensive review with bibliometric insights and future directions. *EuroMed Journal of Business*. Advance online publication. <https://doi.org/10.1108/EMJB-11-2024-0307>
- Lee, Y. J. (2023). Language learning affordances of Instagram and TikTok. *Innovation in Language Learning and Teaching*, 17(2), 408-423. <https://doi.org/10.1080/17501229.2022.2051517>
- Li, Y., Li, X., & Li, X. (2024). Simulation of e-learning in English learner writing teaching based on natural language processing and machine learning algorithms. *Entertainment Computing*, 50, 100690. <https://doi.org/10.1016/j.entcom.2024.100690>
- Lin, C. J., & Mubarak, H. (2021). Learning analytics for investigating the mind map-guided AI chatbot approach in an EFL flipped speaking classroom. *Journal of Educational Technology & Society*, 24(4), 16-35.
- Liu, Y. D., Jing, Y. H., Li, J., Dai, J., Hu, Z. B., & Wang, C. L. (2025). Application of AI in engineering education: A bibliometric study. *Review of Education*, 13(1), e70044. <https://doi.org/10.1002/rev3.70044>
- Luke, K. F., Mary, A., Andrew, T., Aaron, G., & Zelinda, S. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of chatbot and human task partners. *Computers in Human Behavior*, 75, 461-468. <https://doi.org/10.1016/j.chb.2017.05.045>
- Maxwell, J. W. (2006). Resituating constructionism. In J. Weiss, J. Nolan, J. Hunsinger, & P. Trifonas (Eds.), *The international handbook of virtual learning environments* (pp. 279-298). Springer.
- Miao, J., Tian, X., Zhang, L., Zhang, B., & Ma, L. (2025). An analysis of the empirical research on SL/FL teaching and learning based on positive psychology: A systematic literature review. *BMC Psychology*, 13(1), 319. <https://doi.org/10.1186/s40359-025-02576-5>
- Morris, S., O'Reilly, G., & Nayyar, J. (2021). Classroom-based peer interventions targeting autism ignorance, prejudice and/or discrimination: A systematic PRISMA review. *International Journal of Inclusive Education*. Advance online publication. <https://doi.org/10.1080/13603116.2021.1900421>
- Oliveira, J. A., Eskandar, K., Kar, E., de Oliveira, F. R., & da Silva, A. L. (2024). Understanding AI's role in endometriosis patient education and evaluating its information and accuracy: Systematic review. *JMIR AI*, 3, e64593. <https://doi.org/10.2196/64593>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Journal of Clinical Epidemiology*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Rahman, A., Raj, A., Tomy, P., & Hameed, M. S. (2024). A comprehensive bibliometric and content analysis of artificial intelligence in language learning: Tracing between the years 2017 and 2023. *Artificial Intelligence Review*, 57(4), 107. <https://doi.org/10.1007/s10462-024-10729-y>
- Rojas-Sanchez, M. A., Palos-Sanchez, P. R., & Folgado-Fernandez, J. A. (2022). Systematic literature review and bibliometric analysis on virtual reality and education. *Education and Information Technologies*, 28(1), 155-192. <https://doi.org/10.1007/s10639-022-11167-5>
- Sankaran, G., Palomino, M. A., Knahl, M., & Siestrup, G. (2022). A modeling approach for measuring the performance of a human-AI collaborative process. *Applied Sciences*, 12(22), 11642. <https://doi.org/10.3390/app122211642>
- Sherman, B. J., Teemant, A., & Prada, J. (2025). Teachers as the nexus of theory and practice with multilingual learners: A teacher's journey of becoming. *System*, 131, 103664. <https://doi.org/10.1016/j.system.2024.103664>
- Sit, C., Srinivasan, R., Amlani, A., Muthuswamy, K., Azam, A., Monzon, L., & Poon, D. S. (2020). Attitudes and perceptions of UK medical students toward artificial intelligence and radiology: A multicenter survey. *Insights Into Imaging*, 11(1), 14. <https://doi.org/10.1186/s13244-019-0830-7>
- Tai, T. Y., & Chen, H. H. J. (2020). The impact of Google Assistant on adolescent EFL learners' willingness to communicate. *Interactive Learning Environments*, 31(3), 1485-1502. <https://doi.org/10.1080/10494820.2020.1841801>
- Terzieva, V., Ivanova, T., Ivanova, M., Ilchev, S., Djambazova, E., & Petrov, I. (2025). Intelligent educational environments: Recent trends, modeling, and applications. *Applied Sciences*, 15*(7), 3800. <https://doi.org/10.3390/app15073800>

- Wang, X. H., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li, L. L. (2023). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education*, 194, 104703. <https://doi.org/10.1016/j.compedu.2022.104703>
- Wei, J., Marimuthu, K., & Prathik, A. (2022). College music education and teaching based on AI techniques. *Computers & Electrical Engineering*, 100, 107851. <https://doi.org/10.1016/j.compeleceng.2022.107851>
- Wei, L. (2023). Artificial intelligence in language instruction: Impact on English learning achievement, L2 motivation, and self-regulated learning. *Frontiers in Psychology*, 14, 1261955. <https://doi.org/10.3389/fpsyg.2023.1261955>
- Wu, R., & Yu, Z. (2023). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, 55(1), 10-33. <https://doi.org/10.1111/bjet.13334>
- Xie, H., Chu, H. C., Hwang, G. J., & Wang, C. C. (2019). Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Computers & Education*, 140, 103599. <https://doi.org/10.1016/j.compedu.2019.103599>
- Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, 34(3), 327-343. <https://doi.org/10.1017/S0958344022000039>
- Yang, J. W., Choe, K., Lim, J., & Park, J. (2025). Asking questions matters: Comparing the effect of learning through dialog with peers, ChatGPT, and intelligent tutoring system. *Interactive Learning Environments*. Advance online publication. <https://doi.org/10.1080/10494820.2025.2461719>
- Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J.-B., Yuan, J., & Li, Y. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity*, 2021, 8812542. <https://doi.org/10.1155/2021/8812542>

Bio:

Süleyman KASAP, Van Yüzüncü Yıl University, Türkiye

Süleyman KASAP, PhD, is an Associate Professor in the Department of English Language Teaching at Van Yüzüncü Yıl University, Van. With over a decade of secondary-level English teaching before transitioning to higher education, KASAP's current research examines Kurdish-Turkish bilingualism, language attrition, sociolinguistics, psycholinguistics, and the integration of artificial intelligence (AI) in language learning contexts.
suleymankasap@yyu.edu.tr