Computers and Education: Artificial Intelligence 9 (2025) 100446

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

Integrating flipped learning in AI-enhanced language learning: Mapping the effects on metacognitive awareness, writing development, and foreign language learning boredom

Ehsan Namaziandost

Department of English, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran

ARTICLE INFO

Keywords:
Artificial intelligence
Flipped learning
AI-Enhanced flipped learning
Metacognitive awareness
Writing development
Foreign language learning boredom

ABSTRACT

The increasing integration of Artificial Intelligence (AI) within educational contexts is fostering significant pedagogical shifts, with flipped learning concurrently gaining prominence as a dynamic, student-centered instructional model. Despite their individual merits, there remains a paucity of empirical research on the synergistic effects when AI enhances flipped learning, particularly concerning crucial EFL learner variables such as metacognitive awareness, writing development, and foreign language boredom. This study, therefore, aimed to address this gap by examining the impact of an AI-enhanced flipped learning environment. To this end, a quasi-experimental design was implemented involving 70 intermediate Iranian EFL learners, selected through convenience sampling and subsequently assigned to either an experimental group, which experienced AI-enhanced flipped instruction, or a control group that received AI-enhanced instruction without the flipped component. The required data were gathered via validated questionnaires and pre- and post-intervention writing tasks. Following this, a one-way ANCOVA was employed for data analysis to ascertain inter-group differences while controlling for pre-existing variations. The results revealed that AI-enhanced flipped learning significantly improved learners' metacognitive awareness. Furthermore, the findings indicated that this pedagogical approach significantly promoted writing development among students in the experimental group. Additionally, the outcomes showed a substantial reduction in foreign language learning boredom for learners engaged in the AI-enhanced flipped classroom. These findings lend support to the potential of strategically embedding AI within flipped learning models to cultivate more autonomous, engaged, and effective language learning experiences.

1. Introduction

The integration of Artificial Intelligence (AI) into educational settings, particularly in second language (L2) education, has marked a significant paradigm shift, promising to revolutionize traditional teaching and learning methodologies. The development of AI in L2 education, as documented by Warschauer and Healey (1998), began with early Computer-Assisted Language Learning (CALL) programs, which primarily offered drill-and-practice exercises. Over time, advancements in natural language processing, machine learning, and adaptive technologies have propelled AI from a supplementary tool to a more integral component capable of providing personalized feedback, interactive conversational practice, and tailored learning pathways (Shadiev et al., 2024a; Zawacki-Richter et al., 2019). This burgeoning interest, as noted by scholars such as Bibauw et al. (2019) and Golonka et al. (2014), stems from AI's potential to cater to individual learner needs, offer immediate and scalable support, and create more engaging learning environments, thus addressing some of the persistent challenges in L2 acquisition.

Concurrently, innovative pedagogical approaches such as flipped learning have gained traction, especially within English as a Foreign Language (EFL) education. Bergmann and Sams (2012) famously conceptualized flipped learning as a model that fundamentally reconfigures the learning process by moving direct instruction from the group learning space to the individual learning space, typically via instructor-created videos or curated online resources. In this model, class time is repurposed for active, collaborative, and higher-order thinking tasks (Akçayır & Akçayır, 2018; Lee & Wallace, 2018; Vitta & Al-Hoorie, 2023). The synergy between AI and flipped learning has recently given rise to AI-enhanced flipped learning (AIFL) models. In such models, AI tools can further personalize the pre-class learning materials—for instance, through adaptive learning platforms or intelligent tutoring systems—while also supporting in-class activities with

E-mail address: e.namazi75@yahoo.com.

https://doi.org/10.1016/j.caeai.2025.100446
Received 14 December 2024; Received in revised form 7 July 2025; Accepted 12 July 2025
Available online 15 July 2025

AI-driven feedback tools or conversational agents; this optimizes both phases of the flipped classroom (Arslan, 2020; Ray & Sikdar, 2024; Uchiyama et al., 2023; Wang & Reynolds, 2024).

Despite the theoretical advantages and increasing implementation of AI-enhanced learning environments, it is crucial to empirically investigate their impact on multifaceted learner outcomes. Specifically, exploring the effects on EFL learners' metacognitive awareness, their ability to understand and control their own learning processes, a concept extensively discussed by Flavell (1979) and later elaborated by Wenden (1998), is paramount, as enhanced metacognition is strongly linked to improved learning autonomy and success. Furthermore, writing development, a complex skill requiring iterative practice and constructive feedback, stands to benefit significantly from AI tools that can offer scaffolded support and detailed analysis (Ranalli, 2021; Shadiev & Feng, 2024). Equally important is the affective domain, particularly foreign language boredom, which, as Pawlak et al. (2025) has highlighted, can considerably impede motivation and persistence in language learning. While the potential of AIFL to address these areas seems promising, the combined influence of this pedagogical approach on EFL learners' metacognitive awareness, writing proficiency, and boredom levels has, to date, remained largely under-explored in empirical research.

The significance of investigating these variables is particularly pronounced within the EFL context of Iran. As Rezai and Goodarzi (2025) and Sadeghi and Richards (2016) observe, Iranian EFL learners often face challenges such as large class sizes, limited opportunities for authentic language use outside the classroom, and a curriculum that may not always cater to individual learning paces or affective needs. AIFL environments could offer a viable means to provide more individualized attention, offer diverse and engaging learning materials, and foster more active learning experiences, potentially mitigating some of these contextual challenges. Understanding the effects of this innovative approach can consequently provide valuable insights for Iranian educators, curriculum developers, and policymakers seeking to leverage technology effectively to improve EFL learning outcomes and learner engagement in this specific educational landscape. This study, therefore, aims to address the identified gap by examining the impact of AIFL on EFL learners' metacognitive awareness, writing development, and foreign language boredom within the Iranian context.

2. Literature review

2.1. AI-enhanced flipped learning

AI has garnered considerable attention, particularly within L2 education. As Zawacki-Richter et al. (2019) and Chiu et al. (2023) define it, AI in this context refers to computational systems designed to perform tasks typically requiring human intelligence, such as understanding natural language, recognizing speech, making decisions, and adapting to user needs. This burgeoning interest in AI, as Rezai, Namaziandost, and Hwang (2024) and Hardaker and Glenn (2025) note, stems largely from its capacity to address longstanding pedagogical challenges, including the provision of personalized learning experiences and immediate, individualized feedback, which are often difficult to scale in traditional classroom environments. Accordingly, AI has been highly welcomed in L2 education because it offers outstanding benefits for learners; for instance, it can provide tailored learning pathways catering to diverse proficiency levels and learning paces, offer extensive opportunities for interactive practice without fear of judgment, and deliver instant corrective feedback, which fosters learner autonomy and potentially enhances motivation (Rezai, Namaziandost, & Hwang, 2024; Saihi et al., 2024).

Building upon these technological advancements and established pedagogical models, AIFL has emerged as an innovative instructional approach. Flipped learning itself, which Bergmann and Sams (2012) and Akçayır and Akçayır (2018) characterize as the inversion of traditional teaching where direct instruction occurs outside class time and in-class time is dedicated to application and practice, provides a fertile ground for AI integration. The synergy materializes when AI tools are strategically embedded within the flipped learning framework; for example, as Emerson (2024) highlight, AI can power adaptive learning platforms for pre-class content delivery, facilitate intelligent tutoring systems for guided practice, or offer automated feedback on assignments, thus enriching both the out-of-class and in-class learning phases (Phanwiriyarat et al., 2025). This thoughtful combination transforms the conventional flipped model into a more dynamic, responsive, and data-driven AIFL environment.

The important advantages of AIFL are multifaceted and hold significant promise for positively affecting L2 learning. For instance, Uchiyama et al. (2023) notes that by leveraging AI, the pre-class component can transcend simple video lectures, offering interactive, adaptive materials that diagnose and address individual learner weaknesses before they enter the classroom. Subsequently, the in-class experience, guided by insights from AI-generated analytics, allows educators to orchestrate more targeted, collaborative, and engaging activities that focus on higher-order thinking and communicative competence (Huesca et al., 2024; Le, 2024). This refined approach can foster deeper engagement, improve specific language skills through personalized practice, and enhance metacognitive development by making learning processes more transparent and feedback more actionable. Moreover, by tailoring content and pacing, AIFL has the potential to sustain learner interest and mitigate factors such as learning boredom, which can be detrimental to sustained language acquisition efforts (Babu et al., 2025; Ray & Sikdar, 2024).

2.2. Metacognitive awareness

In the domain of L2 education, metacognitive awareness (MA) has increasingly been recognized as a critical factor influencing learning success. Broadly defined for this context, MA refers to learners' conscious understanding and volitional control over their own cognitive processes during language learning activities (Flavell, 1979; Wenden, 1998). This encompasses both metacognitive knowledge, what learners know about themselves as learners, the nature of language learning tasks, and the strategies available to them, and metacognitive regulation, which involves the active processes of planning, monitoring, and evaluating their learning (Anderson, 2012; Teng et al., 2022). For instance, metacognitive knowledge might involve a student understanding that they learn vocabulary best through spaced repetition (person knowledge), that writing an argumentative essay requires specific structural elements (task knowledge), or that outlining before writing is a useful approach (strategy knowledge). Subsequently, according to Anderson (2012), metacognitive regulation would involve them actively planning to use spaced repetition, monitoring their essay structure as they write, and evaluating the effectiveness of their outlining strategy.

Consequently, the cultivation of MA is particularly vital for EFL learners, as it empowers them to become more autonomous and strategic in their learning endeavors. Learners with higher MA are often better equipped to navigate the complexities of language acquisition; they tend to employ more effective learning strategies, demonstrate greater resilience when encountering difficulties, and exhibit enhanced performance in demanding tasks such as academic writing (Lee & Mak, 2018; Shadiev et al., 2024b; Teng et al., 2022). Indeed, research indicates that EFL learners who utilize metacognitive strategies often outperform their peers and report higher levels of self-efficacy, particularly in productive skills like writing, which demand the concurrent management of grammar, syntax, coherence, and audience awareness (Qin et al., 2022; Teng & Qin, 2024). By reflecting on their learning processes and adjusting ineffective strategies, students can significantly improve their language proficiency and reduce cognitive overload, thereby fostering more efficient and effective learning outcomes (Sun et al., 2018).

MA may be affected by AIFL in different ways. Flipped learning inherently creates opportunities for MA development by requiring students to engage in self-directed pre-class preparation, thus promoting planning and self-monitoring (Sun & Wu, 2016; Vidergor, 2023). The integration of AI further amplifies these opportunities; AI tools can offer personalized learning pathways, immediate and targeted feedback, and adaptive scaffolding, all of which can prompt learners to reflect more deeply on their understanding, monitor their progress more accurately, and evaluate the effectiveness of their learning strategies more critically (Chiu et al., 2023; Roll & Wylie, 2016). AI-powered writing assistants, As noted by Rezai, Soyoof, and Reynolds (2024), can provide feedback on grammatical errors or stylistic choices, encouraging learners to evaluate their writing and plan revisions, while adaptive learning systems can adjust task difficulty based on performance, prompting learners to monitor their comprehension and adjust their learning tactics accordingly. Therefore, it is posited that an AIFL environment may significantly contribute to the enhancement of MA in EFL learners by providing structured support for the reflective and regulatory processes central to metacognition.

2.3. Writing development

Writing development was the other construct explored in this study. For the purposes of L2 education, writing development can be defined as the progressive and multifaceted acquisition of skills enabling learners to construct coherent, contextually appropriate, and communicatively effective texts in a language other than their native tongue (Hyland, 2019; Silva & Leki, 2004). L2 writing skills for EFL learners are of paramount importance because they act as critical gateways to higher education, professional advancement, and broader global communication (Badger & White, 2000; Dong, 2024; Weigle, 2002). Consequently, pedagogical approaches to fostering these skills have evolved substantially, transitioning from an early focus on the final written product to encompass more dynamic process-oriented strategies, genre-based instruction emphasizing textual conventions, and, more recently, sociocultural perspectives that view writing as a situated social practice (Ferris & Hedgcock, 2023; Hyland, 2016).

AIFL may influence writing development in EFL learners. AIFL integrates AI-powered tools within the flipped classroom model, offering a novel framework that may reshape how writing is taught (Wang & Tian, 2025; Zawacki-Richter et al., 2019). In this pedagogical model, learners engage with AI-driven instructional content, such as automated writing evaluation systems or intelligent tutoring platforms outside of class, freeing synchronous class time for interactive writing activities and personalized teacher feedback (Casal & Bikowski, 2019; Hardaker & Glenn, 2025; Lawan et al., 2023). AI tools provide immediate, individualized feedback on grammar, mechanics, and style, helping learners identify areas for improvement and refine their texts iteratively. As Barrot (2023) and Chapelle and Sauro (2017) note, this process can foster linguistic accuracy, enhance rhetorical awareness, and cultivate deeper metacognitive engagement with writing, ultimately supporting overall writing development in EFL contexts.

2.4. Foreign language learning boredom

Foreign language learning boredom (FLB) is the last construct explored in this study. FLB can be defined for L2 education as an unpleasant and deactivating academic emotion characterized by a lack of interest, low arousal, and a perception of learning tasks or the overall learning environment as monotonous, repetitive, or lacking in personal relevance (Pawlak, Kruk, et al., 2020). To illuminate its underpinnings, particularly among EFL learners, Pekrun's (2006) Control-Value Theory (CVT) offers a robust framework. Pekrun et al. (2010) further explain that, according to CVT, boredom arises when individuals perceive low control over an activity and assign low value to it. Consequently, in an EFL context, if learners feel they have little agency in their learning process (e.g., through prescribed, unvaried tasks) and fail to see the utility or intrinsic appeal of language activities, they are more susceptible to experiencing boredom.

The importance of understanding and addressing FLB within EFL learning environments is significant. Sustained boredom is not merely a transient feeling of disinterest; rather, as Li and Dewaele (2021) have highlighted, it has been consistently linked to reduced cognitive engagement, diminished intrinsic and extrinsic motivation, and more superficial processing of linguistic input. Furthermore, EFL learners experiencing high levels of boredom may exhibit decreased participation in classroom activities, manifest greater task avoidance, and ultimately demonstrate poorer learning outcomes, potentially leading to lower proficiency levels or even attrition from language programs (Pawlak, Kruk, et al., 2020). For these reasons, identifying pedagogical strategies that can mitigate FLB is paramount for fostering more effective and engaging language learning experiences.

Given these considerations, the integration of AIFL into EFL classrooms may play a significant role in mitigating FLB by addressing key factors such as learner engagement, autonomy, and perceived task value. This pedagogical model strategically incorporates AI-powered tools into both pre-class and in-class phases, offering personalized learning pathways and adaptive content delivery (Dan et al., 2023; Phanwiriyarat et al., 2025). By tailoring pre-class activities to individual proficiency levels and interests, AIFL can increase learners' sense of control (autonomy) and relevance (value)—two critical dimensions linked to reduced FLB (Lo & Hew, 2023; Pawlak, Kruk, et al., 2020). Furthermore, interactive AI tools, such as chatbots and gamified exercises, can make pre-class learning more dynamic, potentially decreasing the monotony often associated with traditional homework.

In this part, we critically review a range of studies to lay ground for the current research. Early investigations, such as He et al. (2016), explored the broader academic impact of flipped instruction in a large undergraduate chemistry course, finding positive effects on students' out-of-class study time, exam performance, motivation, and perceptions. This foundational work underscored the potential of flipping the classroom to enhance student engagement and outcomes. Building on this, Sergis et al. (2017) further substantiated the model's value by demonstrating its positive influence on students' overall learning experiences. Within the specific domain of EFL education, research has progressively validated the flipped model's efficacy across various skills and contexts. For instance, Qiu and Luo (2022) employed a quasi-experimental design in China, confirming that flipped listening instruction significantly improved EFL students' listening performance while concurrently reducing their anxiety levels. Similarly, focusing on productive skills, Hashemifardnia et al. (2021) provided evidence from the Iranian EFL context, showing that flipped instruction positively impacted the speaking complexity, accuracy, and fluency of intermediate learners.

2.5. Related studies in literature

More recent studies have continued to affirm and expand upon these findings, often highlighting the nuanced benefits of the flipped classroom in diverse EFL settings. Torres-Martín et al. (2022), for example, not only demonstrated the positive application of the flipped classroom for improving academic performance among EFL learners but also noted its potential for long-lasting effects. Concurrently, Wagner and Urhahne (2021) highlighted the beneficial outcomes of flipped classroom instruction within EFL secondary education, reinforcing its applicability across different educational levels. The affective and cognitive dimensions have also received attention; Aksoy and Tulgar (2024), for instance, uncovered the positive impacts of teaching English via the Flipped Classroom Model (FCM) on EFL students' motivation and autonomy. Furthermore, in a period marked by educational disruption, Samadi et al. (2024) showed the effects of flipped classroom instruction on enhancing self-regulated learning and higher-order thinking skills among EFL learners during the COVID-19 pandemic. Echoing these

sentiments, Yan et al. (2024) illustrated the positive influence of a flipped EFL course on the overall L2 development of college students. As the integration of technology in education deepens, studies have begun to explore the synergy between AI and flipped learning. Diwanji et al. (2018) focused on leveraging AI and analytics to enhance student preparation for flipped classrooms, aiming to improve engagement monitoring. Building on this, Ray and Sikdar (2024) posited AI's revolutionary potential in reshaping flipped learning through personalized paths and intelligent tutoring, while Uchiyama et al. (2023) demonstrated a practical application by developing an LLM-based system for immediate feedback during flipped classroom preparation. Shifting focus slightly but remaining pertinent, Rahimi et al. (2025, assuming publication as projected) aimed to uncover how personalized L2 motivation and self-regulation influence ChatGPT-assisted language learning, highlighting the growing interest in AI's role in fostering learner-centric educational experiences.

While the aforementioned studies collectively underscore the multifaceted benefits of flipped learning, and increasingly, the potential of AI to augment educational practices, a discernible gap persists, particularly within the Iranian EFL context. Although Hashemifardnia et al. (2021) explored flipped learning's impact on speaking skills in Iran, and other studies (e.g., Samadi et al., 2024) have touched upon cognitive skills like self-regulation, there has been limited investigation into the combined effects of the AIFL environment on a broader spectrum of crucial learner variables. In particular, the synergistic impact of this technologically advanced model on EFL learners' metacognitive awareness, their writing development, and their experience of foreign language boredom remains largely uncharted territory in Iran. Moreover, while AI's role in personalization and feedback is gaining traction (e.g., Ray & Sikdar, 2024; Uchiyama et al., 2023), its specific application within a flipped framework to concurrently address cognitive, skill-based, and affective dimensions for Iranian EFL learners has not been comprehensively examined. Therefore, the present study sought to address this lacuna by systematically investigating the effects of AIFL on these three critical aspects (e.g., metacognitive awareness, writing development, and foreign language boredom) among Iranian EFL learners, aiming to offer a more holistic understanding of its potential in this specific educational setting. Given these points, the following research questions were posed:

1. Does AI-enhanced flipped learning significantly improve EFL learners' metacognitive awareness in EFL learners?
2. Does AI-enhanced flipped learning significantly promote EFL learners' writing development?
3. Does AI-enhanced flipped learning substantially reduce foreign language learning boredom in EFL learners?

3. Method

3.1. Design of the study

The present study employed a quasi-experimental design. This type of research design is characterized by the comparison of groups that receive different treatments or interventions, similar to true experiments; however, it crucially lacks the random assignment of participants to these conditions. As Mackey and Gass (2015) point out, such designs are frequently utilized in educational research where random assignment may be impractical or disrupt existing structures.

For this investigation, a total of 70 male EFL students were selected as participants. These students were subsequently divided into two separate groups to allow for comparison between different instructional approaches during the study period. The selection of a quasi-experimental approach was considered the most appropriate method for this context primarily due to the practical constraints often encountered in real-world educational settings.

3.2. Participants

This research was undertaken at an English language institute in Iran, a common setting for focused language instruction. The study involved a cohort of 70 male English as a Foreign Language (EFL) learners. These participants ranged in age from 16 to 22 years, and their first language (L1) was Persian. A convenience sampling strategy was employed for participant selection. Initially, learners at the institute were considered, and their English language proficiency was formally assessed using the Oxford Quick Placement Test (OQPT). Only those whose scores indicated an intermediate level of English proficiency were invited to participate; this selection aimed to ensure a degree of homogeneity within the sample regarding their baseline language abilities. This specific proficiency level was targeted as it often represents a stage where learners are receptive to new pedagogical approaches and possess foundational skills for tasks like academic writing. The decision to include only male learners was a practical necessity, determined by the gender segregation policies prevalent in educational institutions within this specific regional context.

Prospective participants were initially approached through the institute's administration and their regular language instructors, who provided a general outline of the research aims. Following this initial contact, the researchers met with interested students to offer a comprehensive explanation of the study's objectives, the procedures involved, the time commitment expected, and the voluntary nature of their involvement. Written informed consent was meticulously obtained from each student prior to their inclusion. For participants under the age of 18, consent was additionally secured from their parents or legal guardians Participants were assured of the anonymity of their contributions and the confidentiality of their personal data; all information was coded, and individual identities were protected throughout the data collection and analysis phases. Furthermore, they were explicitly informed of their right to withdraw from the study at any stage without incurring any penalty or negative consequence. Subsequent to providing consent, these 70 learners were then divided into an experimental group (EG, n = 35) and a control group (CG, n = 35) for the purpose of the quasi-experimental design.

3.3. Instruments

In this study, the first instrument employed to gauge participants' initial English language proficiency was the Oxford Quick Placement Test (OQPT) (Syndicate, 2001). The selection of the OQPT was principally driven by its established reputation for providing a rapid yet reasonably comprehensive measure of general English language ability; consequently, as this research aimed to investigate an AIFL intervention, establishing a baseline proficiency was essential for a nuanced understanding of the intervention's impact, controlling for pre-existing language competence. The OQPT, in the version utilized, comprised 70 predominantly multiple-choice items designed to evaluate several key linguistic domains, including Grammar, exemplified by items requiring the correct verb form (e.g., "He __ to the cinema every weekend"), Vocabulary, which might ask for the word best completing a sentence (e.g., "She found the instructions quite __"), and Reading Comprehension, typically involving short passages followed by questions on main ideas (e.g., "What is the main reason the author gives for … ?"). Regarding its scoring procedure, the OQPT generally awards one point per correct answer, with the cumulative raw score referenced against a standardized rubric, often aligning with CEFR levels. Significantly, prior to the main phase of this investigation, the psychometric integrity of the OQPT was carefully evaluated: a pilot test with 32 EFL learners (similar to the main study participants but not part of the final sample) established its reliability, yielding a Cronbach's Alpha coefficient of .88, indicative of high internal consistency. Furthermore, its content validity was affirmed through expert judgment by a panel of three doctoral-level EFL instructors and assessment specialists, who reviewed the items for

relevance, clarity, and appropriateness, resulting in a strong consensus with an average inter-rater agreement on item suitability of .91 (calculated using a Content Validity Index approach).

To assess participants' writing development, a writing test, serving as both a pre-test and post-test, was specifically designed for this research by a panel of three EFL assessment experts; this ensured its direct alignment with the study's core objectives of evaluating improvements in writing proficiency. The task itself required participants to compose a text of approximately 150–200 words on a provided topic familiar to them within a 30-min timeframe, a design choice intended to elicit their best performance by minimizing potential cognitive and cultural load. For the systematic scoring of these written productions, a detailed analytical rubric was employed. This rubric delineated four key criteria: Content (assessing the appropriateness and sufficiency of ideas, 5 points), Organization (evaluating coherence in structure and paragraphing, 5 points), Language Use (focusing on the accuracy of grammar and syntax, 5 points), and Vocabulary (appraising the range and precision of word choice, 5 points), which culminated in a total possible score of 20 points. Consequently, the overall score for each participant's writing was determined by summing the points awarded across these four distinct criteria. Crucially, prior to the main study, the psychometric properties of this writing assessment were rigorously examined. Its content validity was established through an independent review by a separate panel of two experienced EFL educators specializing in writing assessment. These experts confirmed the test's relevance, comprehensiveness, and suitability for measuring the targeted writing skills in line with the study's aims; their collective judgment indicated a high degree of alignment, with an inter-rater agreement on item-objective congruence reaching .90. Furthermore, the reliability of the scoring process using this rubric was ascertained through a pilot test administered to 25 EFL learners, who were comparable to the main study sample but not included in it. The essays from this pilot were independently scored by two trained raters, and the consistency of the rubric's application yielded a Cronbach's alpha coefficient of .89, signifying a high level of reliability in measuring writing abilities before its use in the main investigation.

To explore EFL learners' metacognitive awareness, the third instrument utilized included the Writing Metacognitive Experiences Questionnaire (WMEQ), originally developed by Sun et al. (2021). This questionnaire was selected due to its specific focus on eliciting participants' perceptions of their metacognitive experiences during the writing process, which directly aligns with one of the primary dependent variables of the current investigation. The WMEQ, in its original form, comprises 16 items designed to capture four distinct dimensions of metacognitive experience: Metacognitive Feelings (e.g., affective responses like confidence or uncertainty during writing), Metacognitive Judgments (e.g., evaluations concerning the quality of one's writing process or product), Online Task-Specific Metacognitive Knowledge (e.g., awareness of one's knowledge directly relevant to the current writing task), and Online Task-Specific Metacognitive Strategies (e.g., the planning, monitoring, and regulatory actions undertaken during writing). For the purposes of this study, while the core constructs were maintained, some items were subtly rephrased to better reflect the context of AIFL; for instance, an item assessing strategy use might be adapted to: "When working on writing tasks in the AIFL environment, I consciously considered different ways to organize my ideas before I started writing." Participants responded to these items using a 6-point Likert scale, where '1′ represented "strongly disagree" and '6′ signified "strongly agree." Furthermore, to ensure comprehension and minimize linguistic barriers, the WMEQ was translated into the participants' L1 (Persian) and then back-translated into English by an independent bilingual expert to verify semantic equivalence prior to its administration. It is worth noting, before its deployment in the main study, the adapted WMEQ underwent rigorous psychometric evaluation. A pilot test involving 32 EFL learners, who were similar to the main study participants but not part of the final sample, was conducted to ascertain its reliability; the internal consistency, as measured by Cronbach's Alpha, yielded a coefficient of .87 for the overall scale, indicating a high degree of reliability. Additionally, the content and face validity of the modified instrument were established through expert judgment. A panel of three university professors with significant expertise in metacognition and EFL writing reviewed the questionnaire for clarity, relevance to the study's objectives, and appropriateness for the target population, unanimously confirming its suitability for this research.

The last instrument included Foreign Language Learning Boredom Scale (FLLBS). This instrument was originally developed and validated by Li et al. (2021). The selection of this particular scale was predicated on its robust psychometric properties and its specific design to capture the multifaceted nature of boredom experienced by learners in foreign language contexts. The FLLBS, as conceptualized by its developers, offers a nuanced perspective on learner boredom. While Li et al. (2021) detail specific dimensions, for the scope of this study, we focused on its capacity to measure the overall intensity of boredom. For instance, these dimensions might typically explore aspects such as disengagement (a lack of involvement or interest in learning activities) and apathy (a general indifference or lack of emotion towards the language learning process). In tailoring the instrument to the unique objectives of this investigation, which centered on AIFL, the original items underwent slight modification. This adaptation was crucial to ensure the items accurately reflected the participants' experiences within the specific pedagogical environment under scrutiny. For example, an item addressing general classroom experience was refined to be more context-specific, such as: "I often feel uninterested when completing the AI-driven activities in our flipped English lessons." Responses to all items were gauged using a 5-point Likert scale, ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), allowing for a gradient of affective responses. Recognizing the linguistic background of the participants, the entire instrument was meticulously translated from English into their first language (L1) prior to the main data collection phase. Furthermore, to ascertain the appropriateness and psychometric soundness of the modified FLLBS for the present study's sample and context, rigorous pre-study evaluations were conducted. Specifically, the reliability of the adapted scale was assessed through a pilot test involving a group of 30 EFL learners, distinct from the main study sample but sharing similar characteristics. The results of this pilot test indicated a high degree of internal consistency, yielding a Cronbach's alpha coefficient of .89. Concurrently, the content validity of the modified items was established via expert judgment. To this end, three seasoned researchers in the field of EFL and educational technology reviewed the instrument, evaluating each item for its clarity, relevance to the construct of boredom in the AIFL setting, and overall appropriateness.

3.4. Data collection procedures

The data collection for this study unfolded through a carefully sequenced series of steps, ensuring methodological rigor. Initially, the participants' English language proficiency was formally assessed using the OQPT. This foundational step served to confirm their intermediate-level standing and homogeneity across the groups. Subsequently, to ensure full comprehension and minimize any linguistic ambiguities that might affect participant responses, the last two measures were meticulously translated from English into Persian, the participants' native language (L1). This translation process involved careful back-translation and review to maintain semantic fidelity. Following these preliminary measures, the authors proceeded to recruit two experienced EFL instructors who would facilitate the intervention phases. These instructors then underwent a comprehensive training program specifically designed by the researchers. This training was crucial to equip them thoroughly with the pedagogical principles, technological tools, and procedural details pertinent to both the AIFL model intended for the EG and the AI-enhanced traditional instruction for the CG. Once the instructors were deemed proficient in the respective methodologies, the pre-tests were

administered. All participants, therefore, completed the validated Metacognitive Awareness Questionnaire, a standardized writing proficiency test, and the Foreign Language Learning Boredom Scale to establish baseline measures before the interventions commenced.

The EG participated in the AIFL intervention, a structured 13-session program (60 min per session) designed around flipped learning principles (Bergmann & Sams, 2012) and augmented with AI-driven scaffolding (Luckin, 2018). This intervention was organized into three iterative phases: pre-class preparation, in-class application, and post-class reinforcement, all supported by AI tools to optimize self-regulation, feedback efficiency, and collaborative knowledge construction. The pre-class phase focused on AI-supported independent learning. Students accessed curated multimodal pre-class materials, including interactive videos, annotated readings, and micro-lectures, via an AI-powered LMS (e.g., Moodle with AI plugins or Coursera's AI-guided pathways). Active engagement was facilitated through AI-generated pre-assessments, where students completed diagnostic quizzes adaptive to prior performance (using AI algorithms like A/B testing or Bayesian knowledge tracing) to gauge baseline understanding. Additionally, students submitted short essays or journals (e.g., 300–500 words) through an AI writing assistant (e.g., GPT-4 for formative feedback, Turnitin's AI grammar/style checker, or Grammarly) which provided automated, criterion-referenced feedback on aspects like coherence and argument strength, alongside self-assessment prompts.

In-class sessions transitioned from passive reception to active, problem-based learning (PBL) with AI facilitation. These sessions began with a 10-min warm-up using AI-generated concept reinforcement, such as chatbot-led Q&A or Kahoot! quizzes with adaptive difficulty. The main 40-min segment involved structured collaborative tasks, including peer review with AI-generated rubrics (e.g., Turnitin's Revision Assistant highlighting argument gaps), AI-simulated writing scenarios using tools like NoRedInk or WriteLab for contextualized challenges (e.g., "Adapt this passage for a skeptical audience"), and Socratic seminars where AI moderators (e.g., IBM Watson Debater) dynamically adjusted discussion prompts to deepen critical analysis. Each session concluded with a 10-min instructor-led debrief, integrating AI analytics dashboards (e.g., Google Classroom's progress reports) to highlight class-wide trends.

The post-class phase centered on reinforcement and metacognitive reflection. AI systems (e.g., Squirrel AI) assigned adaptive practice and follow-up tasks based on individual error patterns. Students reviewed their learning analytics via interactive dashboards (e.g., Tableau-powered visualizations) to self-monitor growth. Finally, they submitted metacognitive journals, responding to prompts like "How did AI feedback alter your revision process?", which were analyzed using AI sentiment analysis tools to track engagement shifts. This three-phase, AI-integrated model ensured a closed-loop learning cycle, where AI not only automated feedback but also structured social learning, aligning with constructivist and cognitive load theories.

In contrast, the CG received instruction that, while also AI-enhanced, followed a more traditional and conventional classroom model. For the CG, all instructional activities were conducted entirely within the classroom setting. The teaching methodology centered on direct explanations provided by the instructor, supplemented by AI tools for in-class exercises and practice. Unlike the EG, the CG did not engage in structured pre-class learning activities characteristic of the flipped model, with AI support being primarily embedded within the teacher-led sessions. Finally, upon the conclusion of the 13-session intervention period for both groups, the post-tests were administered. The same instruments employed in the pre-test phase were re-administered to all participants. This final step allowed for the collection of data necessary to evaluate the differential effects of the two instructional approaches on the learners' metacognitive awareness, writing development, and levels of foreign language learning boredom.

3.5. Data analysis procedures

The data collected were meticulously analyzed using SPSS Statistics, Version 26. Initially, descriptive statistics, including means and standard deviations, were computed to provide a foundational overview of the pre-test and post-test scores for EFL learners' metacognitive awareness, writing development, and foreign language boredom, while independent samples t-tests were conducted on pre-test scores to ascertain the initial comparability of the EG and CG. Following this, to rigorously evaluate the impact of the AIFL intervention, a one-way Analysis of Covariance (ANCOVA) was utilized as the primary inferential statistic. This particular procedure was deemed most appropriate as it allowed for a comparison of post-test scores between the groups while statistically controlling for any pre-existing differences identified in their pre-test scores, thereby aiming to isolate more precisely the net effect of the intervention itself. It should be noted that prior to conducting the ANCOVA, essential prerequisite assumptions for this parametric test were carefully examined and confirmed to be met: for instance, the data demonstrated an adequate approximation of normality for each dependent variable within the groups, Levene's test for equality of error variances yielded non-significant results ($p > .05$ for all dependent variables), indicating comparable variance across groups, and the assumption of homogeneity of regression slopes was also upheld, as evidenced by a non-significant interaction effect between the covariate (pre-test scores) and the independent variable ($p > .05$ for each outcome).

4. Results

Before evaluating the treatment effects, preliminary analyses were conducted to ensure baseline equivalence between the EG and CG. Descriptive statistics for the pretest scores are presented in Table 1.

As presented in Table 1, the pretest means across all variables were highly similar between the two groups. The largest observed mean difference was 1.03 points, found in the MA scores (EG: 34.54 vs. CG: 35.57), whereas writing and boredom exhibited even smaller differences at .80 and 1.34 points, respectively. Although these small discrepancies indicated a degree of initial equivalence, they necessitated further statistical confirmation. To substantiate this preliminary observation, independent samples t-tests were performed, and the results are displayed in Table 2.

As shown in Table 2, the results of the statistical analyses indicated no significant differences between the groups on any of the pretest measures (all $p > .05$). Specifically, for metacognitive awareness (MA), $t(68) = -.99$, $p = .32$; for writing, $t(68) = 1.07$, $p = .28$; and for boredom, $t(68) = .91$, $p = .36$. Furthermore, Levene's tests for equality of variances were non-significant (all $p > .05$), confirming the assumption of homogeneity of variance, which supports the use of parametric statistical methods. Descriptive statistics for the posttest metacognitive awareness scores are presented in Table 3.

As presented in Table 3, the experimental group (EG) demonstrated a higher mean score in metacognitive awareness (MA) compared to the CG after the intervention (EG: $M = 46.08$, $SD = 7.72$; CG: $M = 38.42$, $SD = 6.73$). The observed mean difference of 7.66 points suggests that the AIFL approach may have contributed to the development of learners'

Table 1
Results of pretest group statistics.

	Groups	N	Mean	Std. Deviation	Std. Error Mean
MA	EG	35	34.54	4.40	.74
	CG	35	35.57	4.27	.72
Writing	EG	35	13.00	3.24	.54
	CG	35	12.20	2.99	.50
Boredom	EG	35	25.71	6.56	1.10
	CG	35	24.37	5.62	.95

Table 2
Results of independent samples test: Pretest comparison.

	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
MA	.01	.89	−.99	68	.32	−1.02	1.03
			−.99	67.93	.32	−1.02	1.03
Writing	.13	.71	1.07	68	.28	.80	.74
			1.07	67.57	.28	.80	.74
Boredom	.74	.39	.91	68	.36	1.34	1.46
			.91	66.45	.36	1.34	1.46

Table 3
Results of the post-test of metacognitive awareness.

Groups	Mean	Std. Deviation	N
EG	46.08	7.72	35
CG	38.42	6.73	35
Total	42.25	8.15	70

metacognitive abilities. To examine this effect more rigorously while accounting for potential confounding influences of initial ability levels, a one-way ANCOVA was conducted. The results of this analysis are displayed in Table 4.

The analysis of covariance results presented in Table 4 demonstrate a statistically significant between-group difference ($F(1,67) = 19.43$, $p < .001$, partial $\eta^2 = .22$), with the experimental group showing superior metacognitive awareness performance compared to the CG when accounting for baseline scores. This medium effect size, following Cohen's (1988) guidelines, indicates the educational intervention exerted a substantively important influence on the development of metacognitive skills. Of particular note, the analysis revealed no significant covariate effect of pretest scores ($F(1,67) = .14$, $p = .70$), suggesting that participants' initial metacognitive awareness levels were unrelated to their post-intervention outcomes. The descriptive statistics for writing performance across both groups are presented subsequently in Table 5.

The descriptive statistics presented in Table 5 reveal a marked difference in writing performance between the EG ($M = 17.57$, $SD = 1.59$) and CG ($M = 14.85$, $SD = 3.03$). This 2.72-point mean difference, coupled with the experimental group's lower score variability, suggests two important findings: first, that the AIFL intervention was associated with both higher and more consistent writing achievement among participants. Second, these preliminary results indicate the potential efficacy of this pedagogical approach for EFL writing development. To rigorously test these observed differences while controlling for baseline writing competence, we subsequently performed a one-way analysis of covariance, the results of which are detailed in Table 6.

As shown in Table 6, the results of the ANCOVA indicated a statistically significant main effect of group membership on writing improvement, $F(1, 67) = 21.70$, $p < .001$, partial $\eta^2 = .24$, demonstrating that the experimental group outperformed the CG to a significant extent. The effect size, classified as moderate to large, suggests that

Table 4
Results of post-test MA: Between-subjects effects analysis.

Dependent Variable: posttest						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1033.83	2	516.91	9.73	.00	.22
Intercept	1604.86	1	1604.86	30.20	.00	.31
Pretest	7.77	1	7.77	.14	.70	.00
Groups	1032.71	1	1032.71	19.43	.00	.22
Error	3559.54	67	53.12			
Total	129590.00	70				
Corrected Total	4593.37	69				

Table 5
Results of post-test of writing development.

Groups	Mean	Std. Deviation	N
EG	17.57	1.59	35
CG	14.85	3.03	35
Total	16.21	2.76	70

Table 6
Results of post-test writing: Between-subjects effects analysis.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	129.44	2	64.72	10.90	.00	.24
Intercept	1082.40	1	1082.40	182.28	.00	.73
Pretest	.51	1	.51	.08	.76	.00
Groups	128.88	1	128.88	21.70	.00	.24
Error	397.84	67	5.93			
Total	18930.50	70				
Corrected Total	527.28	69				

the intervention had meaningful practical implications. Especially, pretest scores did not significantly predict posttest performance, $F(1, 67) = .08$, $p = .76$, indicating that initial writing ability had no appreciable impact on the treatment outcomes. Posttest boredom scores for both groups are presented in Table 7.

As presented in Table 7, participants in the EG reported lower mean levels of boredom ($M = 17.57$, $SD = 4.71$) compared to those in the CG ($M = 21.28$, $SD = 6.32$). This difference of 3.71 points suggests a notable reduction in boredom related to foreign language learning, indicating that the AIFL approach may contribute to alleviating negative affective experiences in language classrooms. The higher standard deviation observed in the CG reflects greater variability in boredom levels among its members. To assess these differences while accounting for initial boredom levels, a one-way ANCOVA was conducted. The results of this analysis are displayed in Table 8.

As presented in Table 8, the ANCOVA results indicated a statistically significant main effect of group membership on boredom reduction, $F(1, 67) = 10.91$, $p = .002$, partial $\eta^2 = .14$, demonstrating that the experimental group experienced a significantly greater decrease in boredom compared to the CG. The effect size, classified as medium, suggests meaningful practical relevance. In contrast to other outcome measures, pretest boredom scores were found to significantly predict posttest results, $F(1, 67) = 10.23$, $p = .002$, partial $\eta^2 = .13$, indicating that initial levels of boredom partially influenced the outcomes observed after the intervention.

5. Discussion

The primary objective of this research was to investigate the effects of the AIFL environment on EFL learners' metacognitive awareness, their writing development, and their experiences of foreign language boredom. The results derived from a one-way ANCOVA were compelling. Specifically, the findings revealed that engagement with AIFL significantly improved EFL learners' metacognitive awareness. Furthermore, the outcomes indicated a significant promotion in their writing development. Concurrently, the intervention substantially

Table 7
Results of post-test of boredom reduction.

Groups	Mean	Std. Deviation	N
EG	17.57	4.71	35
CG	21.28	6.32	35
Total	19.42	5.84	70

Table 8
Results of post-test boredom: Between-subjects effects analysis.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	521.95	2	260.97	9.51	.00	.22
Intercept	472.02	1	472.02	17.21	.00	.20
Pretest	280.52	1	280.52	10.23	.00	.13
Groups	299.21	1	299.21	10.91	.00	.14
Error	1837.19	67	27.42			
Total	28782.00	70				
Corrected Total	2359.14	69				

reduced levels of foreign language learning boredom among the participating EFL learners.

These findings resonate strongly with a growing body of literature underscoring the pedagogical benefits of flipped classroom models in various educational contexts. For instance, the observed enhancement in learning outcomes, particularly writing development, aligns with earlier work by Thai et al. (2017), who demonstrated superior learning performance in a flipped classroom setup compared to traditional, e-learning, and blended learning environments. More recently, Qiu and Luo (2022) confirmed the positive impact of flipped instruction on Chinese EFL students' listening performance and also noted a reduction in anxiety, a factor often linked to boredom and metacognitive engagement. The improvements in academic performance and metacognitive awareness identified in our study are further corroborated by He et al. (2016), who found that flipped instruction positively influenced exam performance and student motivation in a large undergraduate course. Similarly, focusing on specific language skills, Hashemifardnia et al. (2021) confirmed the positive impacts of flipped instruction on the speaking complexity, accuracy, and fluency of Iranian intermediate EFL learners, while Torres-Martín et al. (2022) highlighted the long-lasting positive effects of the flipped classroom on the academic performance of EFL learners. The current study's demonstration of enhanced cognitive skills, such as improved metacognitive awareness, also parallels the findings of Samadi et al. (2024), who showed that flipped classroom instruction bolstered self-regulated learning and higher-order thinking skills among EFL learners.

The integration of AI within the flipped model, a central feature of our intervention, appears to amplify these established benefits. The observed improvements in metacognitive awareness and writing development, alongside reduced boredom, find further conceptual support when considering recent advancements in AI-enhanced learning. Diwanji et al. (2018), for example, explored how AI and analytics could enhance student preparation in flipped classrooms, thereby fostering deeper engagement, which is a precursor to heightened metacognitive awareness. The revolutionary potential of AI to create personalized learning paths and provide intelligent tutoring, as posited by Ray and Sikdar (2024), likely contributed to the tailored support that facilitated writing development in our study. Moreover, the capacity of AI, such as the LLM-based system developed by Uchiyama et al. (2023) for immediate feedback, can be instrumental in refining writing skills and empowering students to monitor their learning metacognitively. The reduction in boredom and concurrent rise in what could be inferred as increased motivation are also consistent with studies like Aksoy and Tulgar (2024), who found FCM positively impacted EFL students' motivation and autonomy. This connection is further strengthened by Rahimi et al. (2025), whose work on ChatGPT-assisted language learning underscores AI's role in influencing L2 motivation and self-regulation, factors intrinsically linked to sustained engagement and reduced boredom. Thus, the positive outcomes of our AIFL approach are not only consistent with general findings on flipped learning efficacy, such as those by Yan et al. (2024) on L2 development and Wagner and Urhahne (2021) in EFL secondary education, but also align with Sergis et al. (2017) who affirmed the model's positive effect on students' overall learning experiences, suggesting that AI serves as a powerful catalyst in promoting various dimensions of language acquisition within this framework.

The significant improvement in EFL learners' metacognitive awareness observed in this study can be plausibly attributed to several interconnected pedagogical principles inherent in the AIFL model. The interactive and social dimensions, as underscored by Social Constructivist Theory (Vygotsky, 1978), likely played a crucial role. When participants in the experimental group engaged in class discussions, offered peer feedback, and navigated problem-solving exercises, they were essentially externalizing their thought processes and reflecting on their understanding, which are foundational to metacognitive development. This collaborative environment, where learners articulate and refine their comprehension, could have directly fostered their ability to plan, monitor, and evaluate their own learning. Furthermore, the active engagement promoted by the flipped structure (Bergmann & Sams, 2012), where students interact with material pre-class and then apply it, likely necessitated greater self-regulation and reflection on their learning progress, key components of enhanced metacognitive awareness. The AI component, by potentially offering tailored prompts or feedback, may have further scaffolded this reflective process.

The notable promotion of EFL learners' writing development also appears to stem from the synergistic effects of the AIFL. Moving passive content delivery outside the classroom, a core tenet of flipped learning, could have optimized cognitive load (Sweller, 1988), allowing more in-class time to be dedicated to the cognitively demanding task of writing. This active application of knowledge, facilitated by the teacher acting as a guide (Thomas & Philpot, 2012), likely provided more opportunities for focused writing practice than in a traditional setup. The ability for students to personalize their learning journey, perhaps through AI-driven activities or resources as suggested by the potential of AI to tailor content, might have allowed them to address specific weaknesses in their writing more effectively. Moreover, the collaborative activities encouraged by Vygotsky's (1978) framework, such as peer review, could have provided valuable, diverse perspectives on their writing, leading to more robust skill development. The individualized attention and feedback potentially afforded by AI tools may have also allowed students to better understand and rectify their writing errors.

The substantial reduction in foreign language learning boredom among participants in the AIFL group can be largely understood through the lens of Self-Determination Theory (SDT) (Deci & Ryan, 1985). Flipped learning inherently promotes autonomy, as students often have more control over the pace and timing of their engagement with pre-class materials (Little, 2022). This increased sense of control, coupled with the active, participatory nature of in-class sessions where students are not merely passive recipients but constructors of knowledge, likely enhanced intrinsic motivation. When learners feel more competent and autonomous, as SDT suggests, they are less likely to experience negative emotions like boredom. The AI elements might have contributed to this by making pre-class learning more interactive or by providing learners with a sense of progress and mastery through immediate feedback, thus fostering feelings of competence and reducing disengagement.

Ultimately, the superior performance of the experimental group in terms of heightened metacognitive awareness, improved writing skills, and reduced boredom seems to be a consequence of the multifaceted and dynamic learning environment created by the AIFL. The shift towards active learning (Bergmann & Sams, 2012), where students engage with material before class and use in-class time for application and collaboration, likely fostered deeper processing and engagement. The integration of AI could have further personalized this experience, offering tailored support that addressed individual learning needs and promoted self-regulation. This combination of student agency, active participation, opportunities for social learning (Vygotsky, 1978), and potentially optimized cognitive load (Sweller, 1988) appears to have created a more

stimulating and effective learning context, empowering students to take greater ownership of their learning and achieve more positive outcomes compared to a more traditional approach.

6. Conclusions and implications

This study set out to investigate the effects of the AIFL model on EFL learners' metacognitive awareness, their writing development, and their experiences of foreign language learning boredom. The empirical evidence gathered indicates that this pedagogical approach yielded substantial positive outcomes: learners in the experimental group demonstrated a significant improvement in their metacognitive awareness, a notable enhancement in their writing proficiency, and a significant reduction in foreign language learning boredom when compared to their counterparts in the CG. Interpreting these findings, it appears the integration of AI tools within a flipped learning framework cultivated a more active, personalized, and engaging educational milieu. This environment likely fostered increased learner interest in the subject matter, promoted better self-regulatory strategies, and facilitated more effective skill development in writing, as the capacity of AI to offer tailored feedback and adaptive learning pathways, combined with the pre-class exposure inherent in flipped learning, empowered learners to take greater ownership of their learning. Conversely, the traditional pedagogical approach experienced by the CG, characterized by less active participation and fewer opportunities for individualized support, resulted in less pronounced advancements.

The theoretical contributions of this research are noteworthy, lending further empirical support to constructivist learning theories by highlighting how technology-mediated, learner-centered approaches can create conditions conducive to deeper cognitive engagement and knowledge construction. Furthermore, the observed gains in metacognitive awareness align with and extend theories of self-regulated learning, suggesting that AI-enhanced environments can be instrumental in developing learners' abilities to plan, monitor, and evaluate their own learning more effectively, positioning them as more autonomous and strategic learners. This study also contributes to the burgeoning field of technology integration in language education by demonstrating how AI can not only support cognitive development but also positively influence affective domains, such as reducing boredom, which is a critical factor in sustained language learning engagement. In sum, the outcomes underscore the potential of AI-enhanced flipped learning to create more dynamic, engaging, and learner-centric pedagogical spaces that bolster metacognitive reflection and skill acquisition while fostering a more positive affective learning experience.

The findings of this study present several practical implications for various stakeholders invested in enhancing EFL education. For EFL learners themselves, the demonstrated benefits of AIFL, namely, improved metacognitive awareness, stronger writing skills, and reduced boredom signal a clear pathway to more effective and engaging language acquisition. Learners should, therefore, be encouraged to actively participate in such pedagogical models, leveraging the increased autonomy and opportunities for personalized feedback to take greater control of their learning journey. Educators, in turn, are presented with compelling evidence to reconsider traditional teaching paradigms. Specifically, they might explore redesigning their courses to incorporate AIFL, a shift that involves curating or creating suitable pre-class materials and utilizing AI tools to provide tailored support. This transition also necessitates a pedagogical reorientation, moving from a primary role as knowledge transmitters to one as facilitators of learning, orchestrating in-class activities that build upon pre-acquired knowledge through collaborative tasks and critical application.

Beyond the classroom, these results offer valuable insights for curriculum designers and educational technologists. There is a clear impetus to develop and refine AI tools that are not only technologically sophisticated but also pedagogically sound, specifically designed to support the development of metacognitive skills and provide nuanced writing feedback within a flipped learning context. Furthermore, material developers could create resources that explicitly guide learners in using AI feedback effectively and foster self-regulatory learning habits. For educational institutions and policymakers, the study underscores the importance of investing in the necessary technological infrastructure and, crucially, in robust professional development programs. Such programs should equip educators with both the technological competence and the pedagogical adaptability required to successfully implement AIFL models, ultimately fostering more dynamic, personalized, and effective learning environments for EFL students.

6.1. Limitations of the study and suggestions for further research

While this study provides valuable insights into the benefits of AIFL, certain limitations inherent in its design warrant consideration when interpreting the findings. The generalizability of our results is constrained primarily by the sample characteristics, which involved 70 male EFL learners from a single institution in Iran; this specific demographic and geographic context, including gender-segregated educational norms and potentially distinct cultural attitudes towards technology, may not fully represent other learner populations or educational settings. Moreover, the intervention was limited to 13 sessions without a subsequent follow-up to assess the durability of gains, and its exclusive focus on writing skills means its applicability to other language domains like speaking or listening cannot be directly inferred. Finally, the study did not systematically control for potential variations in teacher implementation styles or deeply investigate learners' subjective experiences with AI, such as technical difficulties or the perceived impersonality of feedback, which could have subtly influenced the observed effects.

Acknowledging these constraints, several innovative avenues for future research emerge that could significantly advance our understanding and application of AI in language education. Future investigations could, for instance, employ cross-cultural comparative designs to examine the interplay between AIFL and diverse pedagogical ecosystems, potentially revealing how cultural learning preferences mediate technology acceptance and efficacy. Longitudinal studies are also warranted, not merely to track skill retention, but to explore the evolving nature of learners metacognitive strategies and their affective trajectories (e.g., sustained motivation, shifting perceptions of AI) over more extended academic periods, perhaps even tracking their transition into higher education or professional contexts. Moreover, research could innovatively probe the efficacy of this model in fostering integrated language skills, such as AI-supported project-based learning that combines writing, research, and oral presentation components, or explore the potential of AI to facilitate authentic intercultural communication tasks within a flipped structure. Delving deeper into the human-AI interface, studies might investigate the impact of explainable AI (XAI) in feedback mechanisms on learner trust and the cognitive processing of suggestions, or explore how AI can provide dynamic, adaptive scaffolding that personalizes both pre-class preparation and in-class collaborative tasks based on real-time learner analytics, thus creating truly responsive and evolving learning pathways.

Ethics approval and data availability statement

The study was approved by the Institutional Review Board of Ayatollah Ozma Borujerdi University, under approval ID: ID: abru.2025/an16. Informed consent was obtained from all participants prior to data collection, and their privacy rights were strictly protected throughout the study. The datasets generated and analyzed during the current study are not publicly available due to privacy and ethical considerations, but they are available from the corresponding author on reasonable request.

Declaration of AI-assisted technologies in the writing process

The authors acknowledge the use of ChatGPT to assist in enhancing the clarity, coherence, and overall linguistic quality of the manuscript during its preparation. All outputs generated with the aid of this tool were critically reviewed, revised, and approved by the authors, who assume full responsibility for the final content and conclusions presented in the article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors wish to extend their heartfelt appreciation to the Institutional Review Board (IRB) of Ayatollah Ozma Borujerdi University for their invaluable ethical guidance and formal approval of this research project. Their careful review, insightful feedback, and commitment to upholding ethical standards were fundamental to the design and implementation of this study. The IRB's support not only ensured compliance with ethical protocols but also reinforced our responsibility to safeguard the rights, privacy, and well-being of all participants involved. We are deeply grateful for their dedication and for facilitating a research environment grounded in integrity and academic rigor.

References

Akçayır, G., & Akçayır, M. (2018). The flipped classroom: A review of its advantages and challenges. *Computers & Education, 126*, 334–345. https://doi.org/10.1016/j.compedu.2018.07.021

Aksoy, B. N.Ç., & Tulgar, A. T. (2024). The effects of flipped classroom on EFL students' autonomy and motivation. *GiST Education and Learning Research Journal, 27*. https://doi.org/10.26817/16925777.1727

Anderson, N. J. (2012). Metacognition: Awareness of language learning. In *Psychology for language learning: Insights from research, theory and practice* (pp. 169–187). Palgrave Macmillan UK.

Arslan, A. (2020). A systematic review on flipped learning in teaching English as a foreign or second language. *Journal of Language and Linguistic Studies, 16*(2), 775–797. https://doi.org/10.17263/jlls.759300

Babu, C. S., William, M. P., & Juvin, C. (2025). Artificial intelligence in flipped learning: Advancing quality assurance and personalized education in higher education. In *New technological applications in the flipped learning model* (pp. 169–192). IGI Global Scientific Publishing.

Badger, R., & White, G. (2000). A process genre approach to teaching writing. *ELT Journal, 54*(2), 153–160. https://doi.org/10.1093/elt/54.2.153

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57*, Article 100745. https://doi.org/10.1016/j.asw.2023.100745

Bergmann, J., & Sams, A. (2012). *Flip your classroom: Reach every student in every class every day*. International Society for Technology in Education.

Bibauw, S., François, T., & Desmet, P. (2019). Discussing with a computer to practice a foreign language: Research synthesis and conceptual framework of dialogue-based CALL. *Computer Assisted Language Learning, 32*(8), 827–877. https://doi.org/10.1080/09588221.2018.1535508

Casal, J. E., & Bikowski, D. (2019). A framework for learning with digital resources: Applications for project-based language learning. In *Global perspectives on project-based language learning, teaching, and assessment* (pp. 167–184). Routledge.

Chapelle, C. A., & Sauro, S. (Eds.). (2017). *The handbook of technology and second language teaching and learning*. John Wiley & Sons.

Chiu, T. K., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 4*, Article 100118. https://doi.org/10.1016/j.caeai.2022.100118

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J: Lawrence Erlbaum.

Dan, L., Mohamed, H., & Yue, Z. (2023). A review on the effect of integrating AI-based technology into flipped learning. *Innovative Teaching and Learning Journal, 7*(2), 41–50. https://doi.org/10.11113/itlj.v7.133

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.

Diwanji, P., Hinkelmann, K., & Witschel, H. F. (2018). Enhance classroom preparation for flipped classroom using AI and analytics. *ICEIS*, (1), 477–483.

Dong, L. (2024). "Brave New World" or not?: A mixed-methods study of the relationship between second language writing learners' perceptions of ChatGPT, behaviors of using ChatGPT, and writing proficiency. *Current Psychology, 43*(21), 19481–19495. https://doi.org/10.1007/s12144-024-05728-9

Emerson, N. (2024). AI-enhanced collaborative story writing in the EFL classroom. *Technology in Language Teaching and Learning, 6*(3). https://doi.org/10.29140/tltl.v6n3.1764, 1764-1764.

Ferris, D. R., & Hedgcock, J. S. (2023). *Teaching L2 composition: Purpose, process, and practice*. Routledge.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906–911. https://psycnet.apa.org/doi/10.1037/0003-066X.34.10.906.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning, 27*(1), 70–105. https://doi.org/10.1080/09588221.2012.700315

Hardaker, G., & Glenn, L. E. (2025). Artificial intelligence for personalized learning: A systematic literature review. *The International Journal of Information and Learning Technology, 42*(1), 1–14. https://doi.org/10.1108/IJILT-07-2024-0160

Hashemifardnia, A., Shafiee, S., Rahimi Esfahani, F., & Sepehri, M. (2021). Effects of flipped instruction on Iranian intermediate EFL learners' speaking complexity, accuracy, and fluency. *Cogent Education, 8*(1), 1–19. https://doi.org/10.1080/2331186x.2021.1987375

He, W., Holton, A., Farkas, G., & Warschauer, M. (2016). The effects of flipped instruction on out-of-class study time, exam performance, and student perceptions. *Learning and Instruction, 45*, 61–71. https://doi.org/10.1016/j.learninstruc.2016.07.001

Huesca, G., Martínez-Treviño, Y., Molina-Espinosa, J. M., Sanromán-Calleros, A. R., Martínez-Román, R., Cendejas-Castro, E. A., & Bustos, R. (2024). Effectiveness of using ChatGPT as a tool to strengthen benefits of the flipped learning strategy. *Education Sciences, 14*(6), 660. https://doi.org/10.3390/educsci14060660

Hyland, K. (2016). *Teaching and researching writing* (3rd ed.). Routledge.

Hyland, K. (2019). *Second language writing* (2nd ed.). Cambridge University Press.

Lawan, A. A., Muhammad, B. R., Tahir, A. M., Yarima, K. I., Zakari, A., Abdullahi II, A. H., … Lawan, S. (2023). Modified flipped learning as an approach to mitigate the adverse effects of generative artificial intelligence on education. *Education Journal, 12*(4), 136–143.

Le, M. T. (2024). The influences of AI-enhanced learning on student engagement in English classes. In *Proceedings of IAC in Budapest 2024* (Vol. 69).

Lee, I., & Mak, P. (2018). Metacognition and metacognitive instruction in second language writing classrooms. *Tesol Quarterly, 52*(4), 1085–1097. https://doi.org/10.1002/tesq.436

Lee, G., & Wallace, A. (2018). Flipped learning in the English as a foreign language classroom: Outcomes and perceptions. *Tesol Quarterly, 52*(1), 62–84. https://doi.org/10.1002/tesq.372

Li, C., & Dewaele, J. M. (2021). How classroom environment and general grit predict foreign language classroom anxiety of Chinese EFL students. *Journal for the Psychology of Language Learning, 3*(2), 86–98. https://doi.org/10.52598/jpll/3/2/6

Li, C., Dewaele, J. M., & Hu, Y. (2021). Foreign language learning boredom: Conceptualization and measurement. *Applied Linguistics Review*. https://doi.org/10.1515/applirev-2020-0124

Little, D. (2022). Language learner autonomy: Rethinking language teaching. *Language Teaching, 55*, 64–73. https://doi.org/10.1017/S0261444820000488

Lo, C. K., & Hew, K. F. (2023). A review of integrating AI-based chatbots into flipped learning: New possibilities and challenges. In *Frontiers in education* (Vol. 8). Frontiers Media SA, Article 1175715.

Luckin, R. (2018). Machine learning and human intelligence. *The future of education for the 21st century*. UCL institute of education press.

Mackey, A., & Gass, S. M. (2015). *Second language research: Methodology and design*. Routledge.

Pawlak, M., Derakhshan, A., Mehdizadeh, M., & Kruk, M. (2025). Boredom in online English language classes: Mediating variables and coping strategies. *Language Teaching Research, 29*(2), 509–534. https://doi.org/10.1177/13621688211064944

Pawlak, M., Kruk, M., Zawodniak, J., & Pasikowski, S. (2020a). Investigating factors responsible for boredom in English classes: The case of advanced learners. *System, 91*, Article 102259. https://doi.org/10.1016/j.system.2020.102259

Pawlak, M., Zawodniak, J., & Kruk, M. (2020b). *Boredom in the foreign language classroom: A micro-perspective*. Springer Nature.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology, 102*(3), 531.

Phanwiriyarat, K., Anggoro, K. J., & Chaowanakritsanakul, T. (2025). Exploring AI-powered gamified flipped classroom in an English-speaking course: A case of Duolingo. *Cogent Education, 12*(1), Article 2488545. https://doi.org/10.1080/2331186X.2025.2488545

Qin, C., Zhang, R., & Xiao, Y. (2022). A questionnaire-based validation of metacognitive strategies in writing and their predictive effects on the writing performance of English as foreign language student writers. *Frontiers in Psychology, 13*, Article 1071907. https://doi.org/10.3389/fpsyg.2022.1071907

Qiu, Y., & Luo, W. (2022). Investigation of the effect of flipped listening instruction on the listening performance and listening anxiety of Chinese EFL students. *Frontiers in Psychology, 13*, Article 1043004. https://doi.org/10.3389/fpsyg.2022.1043004

Rahimi, A. R., Sheyhkholeslami, M., & Pour, A. M. (2025). Uncovering personalized L2 motivation and self-regulation in ChatGPT-assisted language learning: A hybrid PLS-

SEM-ANN approach. *Computers in Human Behavior Reports, 17*, Article 100539. https://doi.org/10.1016/j.chbr.2024.100539
Ranalli, J. (2021). L2 student engagement with automated feedback on writing: Potential for learning and issues of trust. *Journal of Second Language Writing, 52*, Article 100816. https://doi.org/10.1016/j.jslw.2021.100816
Ray, S., & Sikdar, D. P. (2024). AI-Driven flipped classroom: Revolutionizing education through digital pedagogy. *Psychology, 7*(2), 169–179, 10.52589/.
Rezai, A., & Goodarzi, A. (2025). Exploring the nexus of informal digital learning of English and online self-regulated learning in EFL university contexts: Longitudinal insights. *Computers in Human Behavior Reports, 18*, Article 100666. https://doi.org/10.1016/j.chbr.2025.100666
Rezai, A., Namaziandost, E., & Hwang, G. J. (2024). How can ChatGPT open promising avenues for L2 development? A phenomenological study involving EFL university students in Iran. *Computers in Human Behavior Reports, 16*, Article 100510. https://doi.org/10.1016/j.chbr.2024.100510
Rezai, A., Soyoof, A., & Reynolds, B. L. (2024b). Disclosing the correlation between using ChatGPT and well-being in EFL learners: Considering the mediating role of emotion regulation. *European Journal of Education, 59*(4), Article e12752. https://doi.org/10.1111/ejed.12752
Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 26*, 582–599. https://doi.org/10.1007/s40593-016-0110-3
Sadeghi, K., & Richards, J. C. (2016). The idea of English in Iran: An example from Urmia. *Journal of Multilingual and Multicultural Development, 37*(4), 419–434. https://doi.org/10.1080/01434632.2015.1080714
Saihi, A., Ben-Daya, M., & Hariga, M. (2024). The moderating role of technology proficiency and academic discipline in AI-chatbot adoption within higher education: Insights from a PLS-SEM analysis. *Education and Information Technologies*, 1–39. https://doi.org/10.1007/s10639-024-13023-0
Samadi, F., Jafarigohar, M., Saeedi, M., Ganji, M., & Khodabandeh, F. (2024). Impact of flipped classroom on EFL learners' self- regulated learning and higher-order thinking skills during the Covid19 pandemic. *Asian-Pacific Journal of Second and Foreign Language Education, 9*(1). https://doi.org/10.1186/s40862-023-00246-w
Sergis, S., Sampson, D. G., & Pelliccione, L. (2017). Investigating the impact of flipped classroom on students' learning experiences: A self-determination theory approach. *Computers in Human Behavior, 78*, 368–378. https://doi.org/10.1016/j.chb.2017.08.011
Shadiev, R., Chen, X., Reynolds, B. L., Song, Y., & Altinay, F. (2024b). Facilitating cognitive development and addressing stereotypes with a cross-cultural learning activity supported by interactive 360-degree video technology. *British Journal of Educational Technology, 55*(6), 2668–2696. https://doi.org/10.1111/bjet.13461
Shadiev, R., & Feng, Y. (2024). Using automated corrective feedback tools in language learning: A review study. *Interactive Learning Environments, 32*(6), 2538–2566. https://doi.org/10.1080/10494820.2022.2153145
Shadiev, R., Reynolds, B. L., & Li, R. (2024a). The use of digital technology for sustainable teaching and learning. *Sustainability, 16*(13), 5353. https://doi.org/10.3390/su16135353
Silva, T., & Leki, I. (2004). Family matters: The influence of applied linguistics and composition studies on second language writing studies—Past, present, and future. *The Modern Language Journal, 88*(1), 1–13. https://doi.org/10.1111/j.0026-7902.2004.00215.x
Sun, J. C. Y., & Wu, Y. T. (2016). Analysis of learning achievement and teacher–student interactions in flipped and conventional classrooms. *International Review of Research in Open and Distance Learning, 17*(1), 79–99. https://doi.org/10.19173/irrodl.v17i1.2116
Sun, Z., Xie, K., & Anderman, L. H. (2018). The role of self-regulated learning in students' success in flipped undergraduate math courses. *The Internet and Higher Education, 36*, 41–53. https://doi.org/10.1016/j.iheduc.2017.09.003
Sun, Q., Zhang, L. J., & Carter, S. (2021). Investigating students' metacognitive experiences: Insights from the English as a foreign language learners' writing metacognitive experiences questionnaire (EFLLWMEQ). *Frontiers in Psychology, 12*(7448812), 1–15. https://doi.org/10.3389/fpsyg.2021.744842
Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285.
Syndicate, uCLE. (2001). *Quick placement test*. Oxford university Press.
Teng, M. F., & Qin, C. (2024). Assessing metacognitive writing strategies and the predictive effects on multimedia writing. *Asia Pacific Journal of Education*, 1–24. https://doi.org/10.1080/02188791.2024.2325132
Teng, M. F., Qin, C., & Wang, C. (2022). Validation of metacognitive academic writing strategies and the predictive effects on academic writing performance in a foreign language context. *Metacognition and Learning, 17*(1), 167–190. https://doi.org/10.1007/s11409-021-09278-4
Thai, N. T. T., De Wever, B., & Valcke, M. (2017). The impact of a flipped classroom design on learning performance in higher education: Looking for the best "blend" of lectures and guiding questions with feedback. *Computers & Education, 107*, 113–126. https://doi.org/10.1016/j.compedu.2017.01.003
Thomas, J. S., & Philpot, T. A. (2012). An inverted teaching model for a mechanics of materials course. In *Proceedings of the ASEE annual conference & exposition*.
Torres-Martín, C., Acal, C., El-Homrani, M., & Mingorance-Estrada, Á. C. (2022). Implementation of the flipped classroom and its longitudinal impact on improving academic performance. *Educational Technology Research & Development, 70*(3), 909–929. https://doi.org/10.1007/s11423-022-10095-y
Uchiyama, S., Umemura, K., & Morita, Y. (2023). *Large language model-based system to provide immediate feedback to students in flipped classroom preparation learning*. ArXiv (Cornell University) https://doi.org/10.48550/arxiv.2307.11388.
Vidergor, H. E. (2023). The effect of teachers' self-innovativeness on accountability, distance learning self-efficacy, and teaching practices. *Computers & Education, 199*, Article 104777. https://doi.org/10.1016/j.compedu.2023.104777
Vitta, J. P., & Al-Hoorie, A. H. (2023). The flipped classroom in second language learning: A meta-analysis. *Language Teaching Research, 27*(5), 1268–1292. https://doi.org/10.1177/1362168820981403
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
Wagner, M., & Urhahne, D. (2021). Disentangling the effects of flipped classroom instruction in EFL secondary education: When is it effective and for whom? *Learning and Instruction, 75*, Article 101490. https://doi.org/10.1016/j.learninstruc.2021.101490
Wang, X., & Reynolds, B. L. (2024). Beyond the books: Exploring factors shaping Chinese English learners' engagement with large language models for vocabulary learning. *Education Sciences, 14*(5), 496. https://doi.org/10.3390/educsci14050496
Wang, C., & Tian, Z. (Eds.). (2025). *Rethinking writing education in the age of generative AI*. Taylor & Francis.
Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching, 31*(2), 57–71. https://doi.org/10.1017/S0261444800012970
Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.
Wenden, A. L. (1998). Metacognitive knowledge and language learning1. *Applied Linguistics, 19*(4), 515–537. https://doi.org/10.1093/applin/19.4.515
Yan, Z., Liu, Q., & Chen, Y. (2024). Examining the impact of flipped learning on second language development: A quasi-experiment in higher education. *Heliyon, 10*(13), Article e33593. https://doi.org/10.1016/j.heliyon.2024.e33593
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*(1), 1–27. https://doi.org/10.1186/s41239-019-0171-0