Integrating large language models into EFL writing instruction: effects on performance, self-regulated learning strategies, and motivation

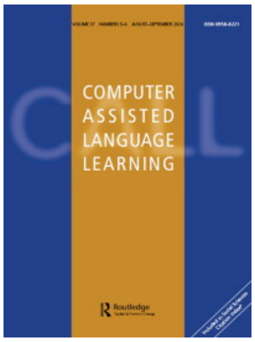
Article *in* Computer Assisted Language Learning · August 2024

DOI: 10.1080/09588221.2024.2389923

5 authors, including:

Ze-Min Liu
East China Normal University

Chuang-Qi Chen
Wenzhou University



Computer Assisted Language Learning

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ncal20

Integrating large language models into EFL writing instruction: effects on performance, self-regulated learning strategies, and motivation

Ze-Min Liu, Gwo-Jen Hwang, Chuang-Qi Chen, Xiang-Dong Chen & Xin-Dong Ye

To cite this article: Ze-Min Liu, Gwo-Jen Hwang, Chuang-Qi Chen, Xiang-Dong Chen & Xin-Dong Ye (12 Aug 2024): Integrating large language models into EFL writing instruction: effects on performance, self-regulated learning strategies, and motivation, Computer Assisted Language Learning, DOI: 10.1080/09588221.2024.2389923

To link to this article: https://doi.org/10.1080/09588221.2024.2389923

COMPUTER ASSISTED LANGUAGE LEARNING
https://doi.org/10.1080/09588221.2024.2389923

Routledge
Taylor & Francis Group

Check for updates

Integrating large language models into EFL writing instruction: effects on performance, self-regulated learning strategies, and motivation

Ze-Min Liu[a], Gwo-Jen Hwang[b], Chuang-Qi Chen[c], Xiang-Dong Chen[d] and Xin-Dong Ye[c]

[a]Department of Educational Information Technology, East China Normal University, Shanghai, China; [b]Graduate Institute of Educational Information and Measurement, National Taichung University of Education, Taichung City, Taiwan; [c]Department of Educational Technology, Wenzhou University, Wenzhou, Zhejiang, China; [d]Faculty of Education, East China Normal University, Shanghai, China

ABSTRACT

This study aimed to investigate the efficacy of utilizing large language models (LLMs) to enhance self-regulated learning (SRL) strategy instruction in English as a Foreign Language (EFL) writing. An LLM-supported Cognitive Academic Language Learning Model (CALLA-LLM) was developed and examined for its potential to improve elementary students' EFL writing performance, SRL strategy use, and writing motivation. In a randomized controlled trial, 65 elementary school students were divided into an experimental group receiving CALLA-LLM instruction and a control group receiving traditional CALLA instruction. Both groups learned SRL strategies over 5 weeks, with data collected pre-intervention, post-intervention, and at a one-month follow-up. Results showed that the CALLA-LLM group made significant improvements in writing performance, SRL strategy use, and writing motivation, maintained most of the gains at follow-up, and significantly outperformed the control group. Findings provide empirical evidence for the efficacy of the CALLA-LLM model in enhancing EFL writing strategy instruction, lending support for integrating AI technologies such as LLMs into English language teaching. Moreover, the study underscores the importance of the "Humans in the Loop" approach, which emphasizes the essential role of human educators in AI-assisted language instruction.

ARTICLE HISTORY
Received 27 January 2024
Accepted 3 August 2024

KEYWORDS
Large language models; AI in education; self regulated learning; English writing; English as a second/foreign language (ESL/EFL)

CONTACT Xin-Dong Ye yxd@wzu.edu.cn Department of Educational Technology, Wenzhou University, Wenzhou, Zhejiang, China 325035.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/09588221.2024.2389923.

1. Introduction

Developing sophisticated English writing skills continues to pose unique challenges for the growing population of English as a Second/Foreign Language (ESL/EFL) learners (Fathi & Rahimi, 2022; Ou et al., 2023). These difficulties, often exacerbated by differences in language structures, cultural norms, and prior educational experiences, have led to a burgeoning focus on fostering ESL/EFL students' development of self-regulated learning (SRL) strategies for writing (Bai & Wang, 2021). SRL refers to the process whereby learners proactively control their learning experiences by setting goals, monitoring progress, and adjusting strategies based on feedback and self-reflection (Zimmerman, 2000). In writing, SRL manifests as using strategies such as goal-setting, planning, monitoring, and revising based on self-evaluation or external feedback to regulate the writing process (Bai & Wang, 2021; Graham & Harris, 2003). Research has shown positive correlations between SRL strategy use and writing proficiency, with students employing these strategies more consistently exhibiting better writing and stronger motivation (Bai, 2015; Teng & Zhang, 2020).

Strategy-based writing instructional interventions are often used to enhance EFL students' SRL strategy use when writing. These interventions typically employ a scaffolding approach, grounded in Social Cognitive Theory (SCT) and Vygotsky's concept of Zone of Proximal Development (ZPD), which involves a gradual transfer of responsibility from the teacher to the learner through modeling, guided practice, structured assistance for independent application, and reflective extension (Chamot, 2005; Vygotskiĭ & Cole, 1978; Zimmerman & Bandura, 1994). However, this approach places high demands on teachers, who often struggle to provide personalized support, real-time assessment, and feedback due to large class sizes, limited teaching time, and scarcity of resources (Finlayson & McCrudden, 2020; Teng, 2022). In response, scaffolding has increasingly transitioned to computer-based environments, including online prompts, tools, and instructional agents (Jiang & Eslami, 2022; Zheng et al., 2023). While these methods have shown effectiveness, they still face challenges in providing sufficient adaptability and personalization to meet individual learner needs and respond to the dynamic nature of the SRL process (Lim et al., 2021, 2023; Zheng et al., 2023). Therefore, innovative approaches are needed to provide more personalized and adaptable scaffolds that support EFL students' SRL strategies and facilitate their writing development.

The emergence of large language models (LLMs) such as GPT-4 has revolutionized the field of language education. LLMs, with their extensive training data and nuanced language faculties, are adept at offering real-time contextualized feedback and fostering adaptive interactions (Kohnke et al., 2023). Recent research suggests that LLMs can be situated within the theoretical frameworks of SCT and ZPD, acting as more knowledgeable others to provide learners with more adaptive and personalized scaffolding (Stojanov, 2023). They can offer real-time feedback and interactions that are sensitive to the individual's learning context, understanding level, and specific challenges (Darvishi et al., 2024). This aligns with the dynamic nature of the SRL process, enabling a more responsive and tailored approach to scaffolding, potentially enhancing students' deep learning and understanding in a way that previous technologies could not (Lin & Chang, 2023).

While the theoretical potential of LLMs in supporting SRL is evident, their practical application and empirical validation in real classroom settings remain largely unexplored. The Cognitive Academic Language Learning Approach (CALLA) is an effective framework for teaching SRL strategies, emphasizing scaffolding methods that progressively develop learners' independent learning skills. Building on this, the present study developed an integrated model (CALLA-LLM) that incorporates LLMs into the CALLA framework, providing a more dynamic and responsive approach to scaffolding. By augmenting traditional methods with personalized, adaptive, and interactive learning experiences, LLMs have the potential to elevate the efficacy of CALLA in equipping EFL learners with SRL strategies.

The present study examined the impact of the CALLA-LLM model on three closely interrelated areas critical to writing success: composition scores, SRL writing strategy use, and writing motivation. Composition scores directly measure writing performance, the ultimate goal of writing instruction. SRL writing strategies, a primary focus of CALLA instruction, are essential for effective writing and have been linked to improved performance (Bai, 2015; Graham & Harris, 2003; Teng & Zhang, 2020). Moreover, social cognitive theory posits that motivational variables like interest and self-efficacy are interrelated and positively correlated with SRL strategy use, highlighting the interdependence of cognitive and motivational factors (Zimmerman, 2000). LLM-based scaffolding, with its adaptive, personalized support and immediate feedback, has the potential to enhance both SRL strategy use and writing motivation, ultimately improving writing performance. By examining these three areas, this study aims to comprehensively evaluate the CALLA-LLM model's effectiveness. The following research questions were explored:

1. Can the CALLA-LLM writing strategy instruction model improve students' English composition scores compared to traditional CALLA instruction?

2. Can the CALLA-LLM model increase students' use of SRL writing strategies (planning, text generation, revising, self-monitoring) compared to traditional CALLA instruction?
3. Can the CALLA-LLM model enhance students' writing motivation, including self-efficacy and interest, compared to traditional CALLA instruction?

2. Literature review

2.1. Self-regulated learning (SRL) strategies in ESL/EFL writing

Self-regulated learning (SRL) strategies are crucial for enhancing ESL/EFL writing proficiency. By integrating SRL principles with process-based writing approaches, learners can more effectively navigate the complex challenges of ESL/EFL writing (Teng & Zhang, 2020; Zhang & Zou, 2022). Learners apply SRL strategies across pre-writing, writing, and post-writing phases, including goal-setting, monitoring, reflection, and feedback-seeking (Fathi & Rahimi, 2022; Graham & Sandmel, 2011). This integration provides learners with a structured and iterative framework that reduces cognitive load (Bai & Wang, 2021), promotes active reflection and feedback-seeking, and enhances writing motivation and performance (Bai, 2015; Sun et al., 2022).

Scaffolding plays a pivotal role in supporting the development of SRL (Chamot, 2005; Zimmerman & Bandura, 1994), and has gradually expanded to include computer-generated support such as prompts and agents embedded in digital learning environments (Jiang & Eslami, 2022; Zheng et al., 2023). However, existing scaffolds often fail to adequately account for the dynamic nature of SRL processes, lacking adaptability and personalization (Lim et al., 2021; Zheng et al., 2023). Van der Graaf et al. (2023) argue that personalized scaffolds should provide support based on learners' actual SRL processes, not merely their learning progress. Furthermore, Lim et al. (2023) emphasize that effective analytics-based personalized scaffolds should combine real-time assessment of learners' SRL with key scaffolding features, such as gradual calibration of support level and content, and fading of support as learners become capable of independent SRL. In summary, to optimize SRL support, future research needs to develop more adaptive and personalized scaffolding approaches.

2.2. Motivation in ESL/EFL writing

Motivation playing a pivotal role in developing ESL/EFL writing skills (Ebadijalal & Moradkhani, 2023). Motivation is a multifaceted construct comprising several interconnected components, such as self-efficacy, interest, perceived task value, and attitudes (Troia et al., 2013). Among these components, self-efficacy and writing interest are two key factors that can influence young writers' abilities (Bai et al., 2022).

Writing self-efficacy, an individual's belief in their ability to complete writing tasks and achieve desired outcomes (Bandura, 1997), is a central component of writing motivation. Numerous studies have demonstrated that writing self-efficacy is an important predictive factor of writing performance, with higher writing self-efficacy associated with better writing outcomes as well as greater persistence, effort, and use of SRL strategies on writing tasks (Rahimi & Fathi, 2022; Teng & Zhang, 2020).

Intrinsic interest in writing is another key motivational factor that influences the quality of students' written work and their engagement with writing tasks. Studies have shown that ESL/EFL learners who are motivated by genuine interest tend to produce higher quality writing and are more likely to persist through challenges (Ebadijalal & Moradkhani, 2023). Renninger and Hidi (2022) emphasized the role of situational interest triggered by specific tasks or contexts in cultivating long-term individual interest in writing. For example, Boo et al. (2015) found that EFL learners who were exposed to engaging writing prompts and provided with autonomy in topic selection exhibited higher levels of intrinsic motivation and writing performance.

The interplay between motivation and SRL strategy use has been a focus of several studies in the ESL/EFL writing context. Zimmerman (2000) posited that students with higher motivation are more likely to engage in SRL strategies, and the use of these strategies can, in turn, enhance motivation. This reciprocal relationship has been supported by empirical research. MacArthur et al. (2016) found that ESL learners who actively employed SRL strategies experienced increased self-efficacy and persistence. Similarly, Bai et al. (2024) found positive correlations between motivation components (interest, self-efficacy, and growth mindset) and SRL strategy use among struggling EFL writers in Hong Kong primary schools. Sun and Wang (2020) demonstrated the significant contributions of both writing self-efficacy and writing SRL strategies to the prediction of writing proficiency among Chinese EFL college students. These studies collectively underscore the close connection between motivation, SRL strategies, and writing performance in ESL/EFL contexts, highlighting the importance of fostering both motivation and SRL strategy use to support learners in overcoming the challenges of ESL/EFL writing. Recognizing these intricate connections, the potential of LLMs to support these factors warrants further exploration.

2.3. Large language models for educational purposes

The integration of Artificial Intelligence (AI) into educational settings is catalyzing paradigmatic shifts in language education (An et al., 2023). LLMs like Claude-2 (Anthropic, 2023) and GPT-4 (OpenAI, 2023) possess unique advantages in this domain owing to their advanced natural language processing capabilities. LLMs can conduct sophisticated dialogues, provide explanatory feedback, and dynamically tailor instruction to student needs (Darvishi et al., 2024). These capabilities offer exciting prospects for writing instruction and supporting students' development of SRL strategies and motivation.

LLMs can facilitate personalized and adaptive learning experiences, which are crucial for fostering SRL strategies. LLMs can analyze language patterns and semantics to deliver highly contextualized, individualized learning experiences (Kohnke et al., 2023) and adjust instructional content based on learners' responses, creating tailored learning pathways (Lin & Chang, 2023). By offering personalized modeling, practice opportunities, and feedback, LLMs can optimize the SRL strategy instruction process, helping learners to gradually internalize and independently apply strategies such as planning, text generation, and revision (Bubeck et al., 2023; Ilieva et al., 2023). Moreover, LLMs can adapt the complexity of their language and feedback to suit individual learners' needs and proficiency levels, aligning with Vygotsky's concept of scaffolding (Vygotskiĭ & Cole, 1978).

In addition to supporting SRL strategies, LLMs have the potential to positively impact learner motivation by promoting active learning. Chatbots, powered by LLMs, can encourage students to take an active role in their learning and provide personalized feedback (Lin & Chang, 2023). This personalized support can lead to increased motivation and engagement in the learning process. A recent empirical study by Silitonga et al. (2023) found that AI chatbot-based learning enhanced students' motivation for learning English writing, confirming the potential of LLMs to improve motivation in language education.

While LLMs harbor transformative potential, they are not intended to supplant the pivotal role of educators. The integration of LLMs into writing instruction should be approached from a "Humans in the Loop" perspective (Cardona et al., 2023), emphasizing the imperative for human intervention, oversight, and interaction to ensure alignment with educational objectives and impart humanistic care essential for fostering motivation, and deep understanding (Selwyn, 2019). The present study's CALLA-LLM model embodies this approach, integrating LLM capabilities within the established CALLA framework to augment, rather than replace, traditional teaching.

3. The LLM-supported cognitive academic language learning model

The Cognitive Academic Language Learning Approach (CALLA), designed by Chamot and O'Malley (1994), provides a comprehensive strategy-based instruction (SBI) framework grounded in SCT and Vygotsky's ZPD construct (Graham & Harris, 2003; Zimmerman & Bandura, 1994). CALLA has proven effective for developing self-regulated learning strategies, offering a structured approach that emphasizes sequencing skill development through demonstrations, guided practice, and gradual transfer of responsibility.

However, to enhance the efficacy of SBI for EFL students, the present study constructed an adapted integrative model termed CALLA-LLM (Figure 1). This model leverages LLMs' capabilities as personalized, adaptive scaffolds within the validated CALLA framework. Integrating these capacities across the stages of CALLA creates an enhanced approach to scaffolding that aims to address some of the challenges of traditional scaffolding approaches, such as difficulties providing adequate personalized feedback and accommodating differences in learning style and pace (Granado-Peinado et al., 2023; Lim et al., 2023).

The CALLA-LLM model interweaves LLM interactions throughout the five key instructional phases of CALLA: preparation, presentation, practice, evaluation, and expansion (Figure 1). LLMs complement teacher-led activities by providing customized explanations, real-time feedback, and adaptive follow-up catering to each student's evolving needs. Specifically:

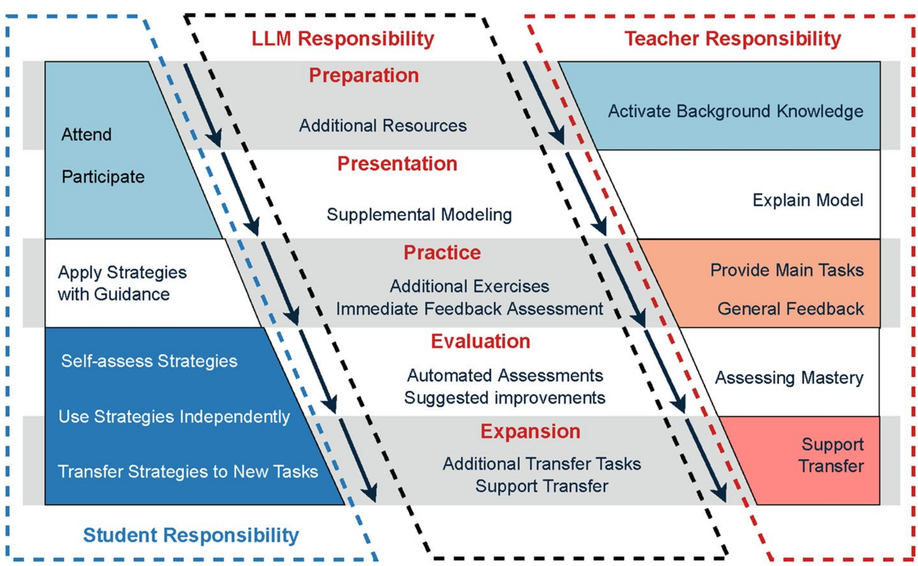


Figure 1. The CALLA-LLM model.

(1) Preparation Phase:

In the preparation phase, teachers activate students' prior knowledge through discussion of their experiences and skills. LLMs provide supplementary textual materials and examples to emp hasize key ideas from the teacher. For instance, the LLM may offer a sample essay with highlighted rhetorical elements to stimulate discussion.

(2) Presentation Phase:

During presentation, teachers introduce target strategies and concepts, modeling potential applications. LLMs augment this instruction with custom examples, supplemental demonstrations, and translations of key ideas into students' native languages when beneficial. LLMs monitor responses to gauge comprehension, adapting explanations when difficulties are detected.

(3) Practice Phase:

In the practice phase, LLMs play a more active role. Teachers assign core practice tasks prompting strategy use. Meanwhile, LLMs generate personalized assignments catering to learners' evolving proficiency levels. During practice, LLMs offer real-time feedback, provide hints, identify misconceptions, and allow resubmission of responses after corrections to support mastery.

(4) Evaluation Phase:

In the evaluation phase, teachers assess overall strategy use, provide holistic feedback, and measure understanding of broader concepts. Simultaneously, LLMs conduct granular analysis of work, quantifying strategy usage, identifying overused/underused strategies, detecting recurring errors, and highlighting areas needing improvement. Students access this analysis to guide further practice.

(5) Expansion Phase:

For the expansion phase, teachers guide generalization of strategies to new contexts, introducing additional writing tasks. Simultaneously, LLMs produce personalized transfer tasks tailored to learners' needs and interests to cement generalization. During these activities, LLMs continue supplying real-time feedback and adaptive guidance prompting strategies.

Across stages, the teacher remains central, providing overall guidance, orchestrating the instructional process, and ensuring course objectives are achieved. Meanwhile, the LLM serves as a dynamic, intelligent scaffold that enhances personalization and the immediacy of feedback catering to each student's ZPD. This balance ensures that responsibility is progressively released to students, empowering them to independently apply and adapt the strategies they have learned. This blended approach aims to elevate the efficacy of CALLA for fostering self-regulated learning.

3.1. Implementation using the CALLA-LLM model

The CALLA-LLM model represents an evolution from the traditional CALLA model that incorporates the dimension of a LLM, introducing a dynamic, real-time interactive layer. This paradigm retains the teacher's role as the core guide, while leveraging the LLM as an adaptive and complementary instructional tool. The LLM is particularly active during the practice, evaluation, and expansion phases, providing real-time feedback and additional exercises or resources tailored to students' proficiency levels.

3.1.1. Modular prompt design

Deploying GPT-4 (or any LLM) within an educational milieu necessitates a careful consideration of prompt construction to maximize model efficacy. Given the recursive nature of SBI pedagogy and its emphasis on progressively relinquishing responsibility to learners, this study adopted a modular hybrid approach (Liu et al., 2022) combining general curriculum prompts with phase-specific prompts to ensure the relevance and specificity of each course phase. Specifically:

1. General Curriculum Prompts: At the start of each lesson, general prompts delineating course objectives, anticipated outcomes, target student grade level, and other meta-information lay the groundwork for LLM interactions.
2. Phase-Specific Prompts: As the curriculum progresses through various phases (preparation, presentation, practice, evaluation, expansion), educators can input or toggle custom prompts consistent with the objectives of each phase. This ensures that feedback and guidance from the LLM are pertinent and specific to the current course phase.
3. Incorporating Dynamic Elements: Students are granted the liberty to interact freely with the LLM during select portions of the course. They can pose specific queries, seek clarifications, or request additional examples.

For detailed illustrations, please refer to the appendix, which showcases the curriculum design for Lesson 8 (see Online Appendix A) under both the CALLA and CALLA-LLM models, complemented by modular prompts for GPT-4 (see Online Appendix B).

3.1.2. API interaction and configuration

Interactions with the LLM are facilitated *via* OpenAI's API. Within the API, the "system" role is designated to define the LLM's behavior, particularly during curriculum prompt setups. Throughout the research process, the system field is dynamically updated to mirror the evolving curriculum stages. This study employed the GPT-4-0314 model and instituted specific API configurations, including:

Temperature: Set to 0.5
Maximum Length: 512
Top P: 0.85
Frequency Penalty: 0.3
Presence Penalty: 0.2

3.1.3. Platform and data privacy

Ensuring seamless interaction with the LLM while protecting data privacy was critical. This research developed a dedicated web application based on open-source code (Hahahumble, 2023). Educators can, in real-time, update the "system" prompts in accordance with the curriculum's progression. Most critically, all sensitive data, encompassing personal configurations and chat logs, are securely stored within the local browser's indexedDB database. Interactions with GPT-4 transpire exclusively through OpenAI's API, thus ensuring rigorous data privacy standards.

4. Methods

4.1. Participants

This study was conducted as part of a summer camp program organized by the sponsoring university. Participants were recruited from several sixth-grade classes of the same elementary school, which was on summer break during the study period. This allowed for a randomized controlled trial (RCT) design, where participants were randomly allocated to either the control or the LLM group, forming two separate classes for the duration of the study. Study protocols were devised through engagement and approval of the school principal and several grade-level teachers. Sixth-grade students were chosen as the target population because they are at a critical stage in the development of their language and writing skills, yet they receive far less attention compared to students in higher education, especially in EFL contexts (Bai et al., 2022). Providing effective SRL strategy interventions at this early stage may have a lasting impact on their EFL writing development.

Inclusion criteria encompassed normal or corrected-to-normal vision, normal cognitive and motor functioning, and no diagnosed health conditions or psychological impairments that could impact study outcomes. Researchers personally reached out to the parents or guardians of the students, presenting and elucidating the research protocol. Parents or guardians who consented to their child's participation signed an informed consent form.

A total of 65 students (37 girls; mean age = 11.4 ± 0.53 years) secured written informed consent from their parents or guardians, ultimately completing the course and three evaluations, and were included in data analysis procedures. Among these, 32 (18 girls; mean age = 11.5 ± 0.57 years) were randomized into the LLM group receiving writing strategy instruction based on the CALLA-LLM model. The remaining 33 (19 girls; mean age = 11.3 ± 0.48 years) were instructed using the traditional CALLA model, serving as a control group.

A priori power analysis was conducted using G*Power 3.1 (Faul et al., 2009) before commencing this study (effect size $f = 0.14$; $\alpha = 0.05$; power = 0.95; correlation between repeated measurements $r = 0.80$), yielding a minimum sample size of 56.

4.2. Procedure

The intervention curriculum was delivered by two experienced English teachers recruited from another elementary school. Prior to the instructional intervention, two educators underwent a 4-week training program, which included bi-weekly seminars lasting an hour each. This training covered SRL strategies, integration of SBI with process writing, and the application of either the CALLA or CALLA-LLM model, ensuring the teachers were well-prepared to deliver the intervention curriculum.

Figure 2 illustrates the experimental procedure. Over 5 weeks, both the LLM and control groups engaged in 10 SBI sessions delivered by the trained teachers, focusing on the same SRL-related writing strategies. The only difference was the instructional model: the LLM group received writing strategy instruction based on the CALLA-LLM model, while the control group adhered to the traditional CALLA model. LLM group student were provided with iPad to access the GPT-4 Web application during the sessions. Assessments were conducted at Baseline (T0), Immediate Posttest (T1), and One-month Follow-up (T2). All measurements were conducted by an independent team blind to intervention details to ensure blinding of students, teachers, and researchers.

All participants and their parents/guardians provided written consent and understood the research design and GPT-4 involvement. Post-research, control group students partook in three condensed CALLA-LLM sessions, emphasizing the key strategies and functionalities of GPT-4, ensuring equitable access to potential intervention benefits. All protocols, procedures, and instructional plans were approved by the research

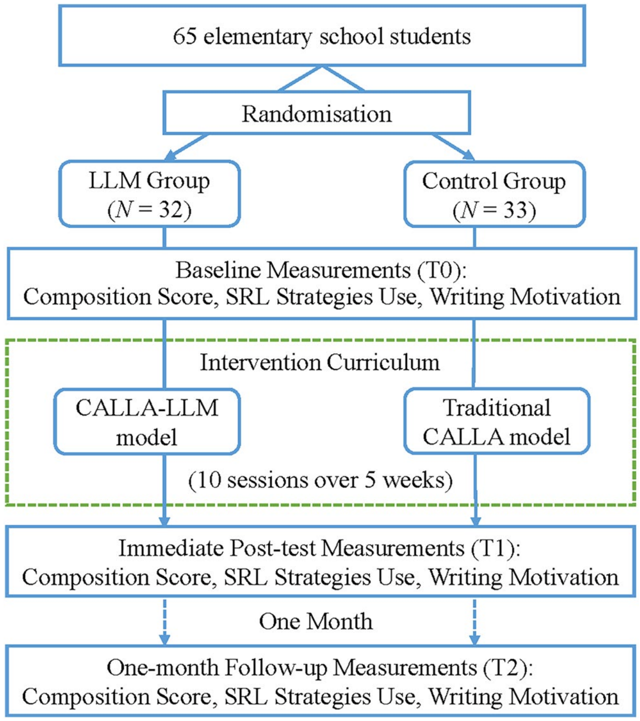


Figure 2. Experimental procedure.

team, elementary school administration, English teachers, and the university's Institutional Review Board, adhering to the Declaration of Helsinki.

4.3. Intervention curriculum

The intervention curriculum was grounded in a process-oriented writing instruction framework (Harris et al., 2010), was based on a 10-lesson, 5-week structure inspired by Bai (2015). Both the LLM and control groups received instruction on the same SRL writing strategies, with identical class durations and teaching objectives. The curriculum progressing through:

1. **The Planning Phase (Lessons 1–4):** Students engaged in goal setting, brainstorming, organizing concepts, and considering audience needs. Planning strategies and self-monitoring techniques were emphasized.
2. **The Drafting Phase (Lessons 5–7):** The focus pivoted to content generation through vocabulary enhancement, syntactic variety, cohesion to improve flow and coherence. Text generation strategies and self-monitoring were critical.

3. **The Revising Phase (Lessons 8–10):** Learners evaluated drafts and refined them via feedback assimilation and structural improvements. Revision strategies and self-regulation was honed. Learners were prompted to integrate strategies from all phases to perfect final pieces.

The key difference between the two groups was the instructional model employed. The LLM group received writing strategy instruction based on the CALLA-LLM model, which integrated LLM support throughout the phases of preparation, presentation, practice, evaluation, and expansion. In contrast, the control group adhered to the traditional CALLA model without LLM integration.

Central to the curriculum were four SRL writing strategy clusters—planning, text generation, revision, and self-monitoring—derived from cognitive process theory (Flower & Hayes, 1981) and empirical research on effective writing instruction (Harris et al., 2010). The curriculum emphasized the synergistic operation of these strategies to facilitate competent writing development (Bai, 2015).

4.4. Measures

4.4.1. Writing competence

Students' composition scores were utilized to measure their writing competence. Three writing tasks were designed for the Baseline, Posttest, and Follow-up assessments, with topics being "My Favorite Holiday," "My School," and "My Favorite Book," respectively. The writing tasks were designed to align with the students' grade level and the curriculum objectives. Each task required students to write a short essay of no less than 70 words, introducing and describing the given topic. Students typically required approximately 30-40 min to complete each task.

Composition scoring employed Hyland's (2003) analytical scoring criteria, recognized for its comprehensive delineation of writing facets. The assessment comprised three segments: sentence structures and vocabulary (max 40 points), organization and coherence (max 20 points), and format and content (max 40 points).

To ensure reliable and objective assessment, two senior primary school educators with extensive experience in English instruction were entrusted with the scoring process. A training session was organized prior to actual assessment to align their understanding of the scoring criteria and calibrate their scoring techniques. Each composition was scored independently by the two raters, with the final score for each composition representing the average of the scores given by the two raters. Inter-rater reliability was robust with Spearman's rho = 0.88, $p < .001$.

4.4.2. SRL writing strategy use

The writing strategies questionnaire used by Bai et al. (2022) was employed to measure participants' use of SRL strategies in EFL writing, encompassing planning strategies (4 items, $\alpha = 0.85$), text generation strategies (4 items, $\alpha = 0.88$), revision strategies (4 items, $\alpha = 0.79$), and self-monitoring strategies (5 items, $\alpha = 0.82$), for a total of 17 items. A 5-point Likert scale was used, with higher scores indicating more frequent use of the respective SRL writing strategy.

4.4.3. Writing motivation

Measurement of writing motivation was adapted from Pintrich and De Groot (1990), comprising nine items, with five used to measure writing self-efficacy and four for writing interest. In the present study, Cronbach's α for the writing self-efficacy subscale was .87 and for the writing interest subscale it was .84. All items used a 5-point Likert scale, with higher scores indicating greater self-efficacy and interest in English writing.

4.5. Statistical analysis

Statistical analysis was performed using Jamovi 2.3 (The Jamovi Project, 2022). An initial exploration of the means and standard deviations of all key variables at each time point was undertaken through descriptive statistics. Normal distribution of all data and residuals was determined through Shapiro-Wilk tests and visual assessment of QQ plots. Homogeneity of variance was ensured by performing Levene's tests. Independent samples *t* tests were then executed at baseline to compare the two groups and establish initial homogeneity. Given the repeated measures nature of the data, linear mixed effects models (LMM) were employed as the primary analyses to examine longitudinal intervention effects on all measurement variables. This approach accounted for fixed effects (group, time, the interaction between group and time) and random effects (participant intercepts), used restricted maximum likelihood (REML) to estimate variance components, and employed Satterthwaite adjustments to calculate degrees of freedom. To further probe effects of time within each group, simple effects analyses were conducted to inspect changes across timepoints in the LLM and control groups separately. Differences were considered statistically significant when probabilities were less than 5% ($p < .05$).

5. Results

Table 1 presents the descriptive statistics for all measurement outcomes. Preliminary observation revealed a salient trend of the LLM

Table 1. Descriptive statistics and time effects analysis.

	Measures, Mean (*SD*)			*F*	*p*
	T0	T1	T2		
Composition scores					
LLM Group	63.05 (9.78)	72.44 (10.74)	74.92 (7.33)	31.9	**< 0.001**
Control Group	62.12 (11.13)	67.35 (8.33)	68.86 (8.40)	10.5	**< 0.001**
Planning strategy					
LLM Group	3.06 (0.51)	3.57 (0.49)	3.52 (0.72)	12.73	**< 0.001**
Control Group	3.11 (0.55)	3.33 (0.52)	3.2 (0.6)	1.93	0.149
Text generation strategy					
LLM Group	3.08 (0.65)	3.49 (0.59)	3.55 (0.67)	12.8	**< 0.001**
Control Group	3.01 (0.64)	3.14 (0.56)	3.08 (0.79)	0.94	0.394
Revision strategy					
LLM Group	3.19 (0.55)	3.66 (0.48)	3.59 (0.67)	12.17	**< 0.001**
Control Group	3.25 (0.57)	3.45 (0.54)	3.30 (0.72)	2.01	0.138
Monitoring strategy					
LLM Group	3.28 (0.45)	3.75 (0.48)	3.64 (0.45)	16.24	**< 0.001**
Control Group	3.34 (0.53)	3.51 (0.48)	3.39 (0.59)	2.08	0.129
Writing self-efficacy					
LLM Group	2.88 (0.46)	3.49 (0.56)	3.23 (0.72)	24.02	**< 0.001**
Control Group	2.92 (0.53)	3.25 (0.65)	3.02 (0.62)	7.41	**< 0.001**
Writing interest					
LLM Group	3.17 (0.52)	3.66 (0.58)	3.53 (0.57)	14.63	**< 0.001**
Control Group	3.11 (0.54)	3.32 (0.52)	3.27 (0.7)	2.98	0.055

Boldface entries indicate statistical significance ($p < 0.05$). SD, standard deviation.

group exhibiting an upward trajectory overall across timepoints, especially post-intervention. Although the control group following traditional instruction also demonstrated gains, they were relatively muted in comparison. Figure 3 depicts the estimated mean trends over time for the variables in both groups, further encapsulating this overall tendency.

5.1. Baseline differences

Pre-intervention assessments confirmed consistency between the LLM and control groups at baseline levels. Specifically, independent samples *t* tests on composition scores, SRL writing strategies, writing self-efficacy, and writing interest at T0 yielded *p*value ranging from 0.620 to 0.746, evidencing no statistically significant differences between the two groups. This initial baseline homogeneity strengthens the validity of subsequent comparative analyses.

5.2. Dynamics of composition scores

As the instructional intervention progressed, the LLM group exhibited a distinctive trajectory in their composition scores. LMM results revealed that, from baseline (T0) to the posttest (T1), the increase in composition scores for the LLM group was estimated to be 4.16 units higher than what was observed for the control group (Table 2). However, this difference was not statistically significant (p = .061). More notably, from the

Table 2. Key outcomes from linear mixed model analysis.

	Effect	Estimate	SE	95% CI	t	p	R^2 Marg	R^2 Cond
Composition scores	(Intercept)	62.12	1.63	(58.92, 65.32)	38.034	**<.001**	0.195	0.640
	Group×Time (T0–T1)[a]	4.16	2.20	(−0.15, 8.48)	1.891	0.061		
	Group×Time (T0–T2)[b]	5.13	2.20	(0.82, 9.45)	2.331	**0.021**		
Planning strategy	(Intercept)	3.11	0.1	(2.92, 3.31)	31.436	**<.001**	0.102	0.458
	Group×Time (T0–T1)	0.3	0.16	(−0.01, 0.60)	1.907	0.059		
	Group×Time (T0–T2)	0.37	0.16	(0.07, 0.67)	2.385	**0.019**		
Text Generation Strategy	(Intercept)	3.01	0.11	(2.78, 3.23)	26.34	**<.001**	0.095	0.654
	Group×Time (T0–T1)	0.28	0.14	(0.00, 0.56)	1.953	0.053		
	Group×Time (T0–T2)	0.39	0.14	(0.11, 0.67)	2.764	**0.007**		
Revision Strategy	(Intercept)	3.25	0.10	(3.05, 3.45)	31.445	**<.001**	0.079	0.555
	Group×Time (T0–T1)	0.27	0.15	(−0.01, 0.56)	1.877	0.063		
	Group×Time (T0–T2)	0.35	0.15	(0.07, 0.64)	2.439	**0.016**		
Monitoring strategy	(Intercept)	3.34	0.09	(3.17, 3.51)	38.40	**<.001**	0.10	0.571
	Group×Time (T0–T1)	0.30	0.12	(0.06, 0.54)	2.470	**0.015**		
	Group×Time (T0–T2)	0.31	0.12	(0.07, 0.55)	2.544	**0.012**		
Writing self-efficacy	(Intercept)	2.92	0.1	(2.72, 3.12)	28.239	**<.001**	0.114	0.685
	Group×Time (T0–T1)	0.29	0.12	(0.04, 0.53)	2.292	**0.024**		
	Group×Time (T0–T2)	0.25	0.12	(0.01, 0.50)	2.033	**0.044**		
Writing Interest	(Intercept)	3.11	0.1	(2.91, 3.30)	30.979	**<.001**	0.101	0.625
	Group×Time (T0–T1)	0.27	0.13	(0.02, 0.53)	2.087	**0.039**		
	Group×Time (T0–T2)	0.19	0.13	(−0.06, 0.45)	1.477	0.142		

R^2 Marg, $R^2_{Marginal}$; R^2 Cond, $R^2_{Conditional}$; T0, Baseline; T1, Immediate Post-test; T2, One-month Follow-up. Boldface entries indicate statistical significance ($p < 0.05$).

[a]The Group×Time (T0-T1) interaction assesses the differential effect of being in the LLM group compared to the control group from T0 to T1.

[b]Similarly, the Group×Time (T0-T2) interaction evaluates the differential effect from T0 to T2.

baseline (T0) to one-month follow-up (T2), the increase in scores for the LLM group surpassed the control group by approximately 5.13 units, a difference that was statistically significant (p = .021). The fixed effects in the model accounted for approximately 19.5% of the variance in composition score differences (R^2_{marginal} = 0.195), while the full model incorporating random effects elucidated approximately 64% of the variance ($R^2_{\text{conditional}}$ = 0.640).

Analyses of time effects further substantiate these observations. Both the LLM group and the control group manifested significant time effects, with the former exhibiting an F-value of 31.9 (p < .001), underscoring a substantial shift in scores. The control group's F-value of 10.5 (p < .001) indicated a positive trajectory, albeit with a notably smaller gradient.

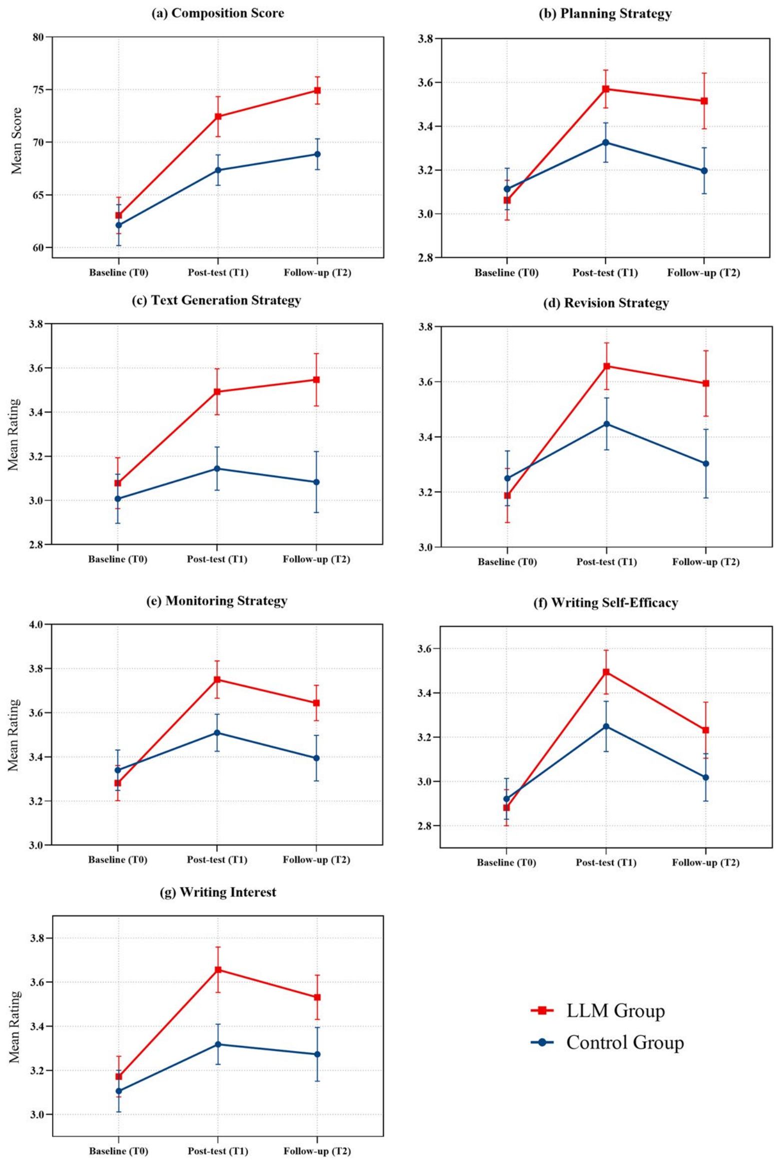


Figure 3. Longitudinal changes in study variables by group.

Overall, the LLM group's composition scores evidenced a statistically significant escalation over time, particularly compared to controls. This divergence was most pronounced at T2, underscoring the sustained advantages of the CALLA-LLM instructional approach.

5.3. Evolution of writing strategies

Similar patterns emerged for writing strategy use. Specifically, from T0 to T1, the increase in self-monitoring strategy ratings for the LLM group was estimated to be 0.3 units higher than the increase observed in the control group, a statistically significant difference (p = .015). In contrast, for the planning (p = .059), text generation (p = .053), and revision (p = .063) strategies, the increase in usage ratings for the LLM group compared to controls was positive but not statistically distinct.

More saliently, when considering change from baseline to one-month follow-up (T0–T2), statistically significant between-group differences emerged for ratings of all four SRL writing strategies, namely planning (p = .019), text generation (p = .007), revision (p = .016), and monitoring (p = .012). The increases in ratings for all SRL writing strategies were estimated to be markedly higher in the LLM group compared to controls, suggesting that CALLA-LLM instruction may confer more lasting impacts on SRL writing strategy acquisition versus traditional approaches.

Analyses of time effects further corroborated this tendency. The LLM group displayed significant time effects for all strategies: planning ($F = 12.73$, $p < .001$); text generation ($F = 12.8$, $p < .001$); revision ($F = 12.17$, $p < .001$); and monitoring ($F = 16.24$, $p < .001$). In contrast, the control group did not exhibit significant time effects for any strategies, with F-values ranging from 0.94 to 2.08 and $p$value all above 0.05. Coupled with the descriptive statistics (Table 1), it is evident that the control group's scores often dropped markedly from T1 to T2, regressing back towards baseline levels, indicating diminishing retention of learned strategies over time. In comparison, the LLM group retained most of the gains from T1 to T2. While some strategies declined slightly, scores remained above baseline.

5.4. Motivation indices: self-efficacy and writing interest

For writing motivation, the LLM group demonstrated significant advantages post-intervention (T0–T1). Specifically, from T0 to T1, the LLM group's increase in ratings for writing self-efficacy was estimated to be approximately 0.29 units higher than that of the control group (p = .024), and the increase in ratings for writing interest was estimated to be approximately 0.27 units higher (p = .039), both reaching statistical significance.

However, when the timeframe extended to one-month follow-up (T0–T2), the advantage of the LLM group in writing motivation seemed to wane. Looking at rating changes from T0–T2, the LLM group only exhibited significantly higher than the control group on writing self-efficacy rating changes (p = .044), but not on writing interest (p = .142).

Analyses of time effects further corroborated these subtle findings. For writing self-efficacy, both groups demonstrated significant time effects, but the effect in the LLM group ($F = 24.02$, $p < .001$) was considerably stronger than in the control group ($F = 7.41$, $p < .001$). In terms of writing interest, the LLM group exhibited a robust time effect ($F = 14.63$, $p < .001$), while controls had a weaker effect, approaching a non-significant effect ($F = 2.98$, $p = .055$).

In summary, while the LLM approach appeared highly effective in terms of immediately enhancing writing self-efficacy and interest post-intervention, the longitudinal data highlighted challenges in sustaining intrinsic motivation such as writing interest over time.

6. Discussion

This study aimed to evaluate the efficacy of incorporating LLMs into SRL strategy instruction for EFL writing. The CALLA-LLM instructional model was developed and compared to the traditional CALLA model to examine its impacts on elementary students' EFL writing performance, strategy use, and writing motivation.

A consistent thread running through the results was the emergence of noticeable divergences between the LLM and control groups over time. The LLM group exhibited substantial and sustained improvements across various facets of writing, suggesting that the CALLA-LLM approach may have imparted meaningful and lasting benefits. In contrast, the control group showed more modest and transient gains from traditional instruction alone. These findings furnish preliminary evidence that integrating LLM support into language instruction can provide ongoing advantages.

6.1. Effects on writing performance and SRL strategies use

The observed incremental improvements in writing quality and SRL strategies use align well with the principles of SRL theory. Self-regulation is not an inherent trait but an acquired skill emanating from structured training environments (Zimmerman, 2000). In this study, the CALLA-LLM model appears to have catalyzed this transition. By offering individualized guidance within each student's ZPD, the LLM acted as a "more knowledgeable other" (Vygotskiĭ & Cole, 1978), likely promoting a deeper understanding and more effective application of the taught strategies, leading to enhanced writing performance (Bai, 2015; Graham & Harris, 2003).

The CALLA-LLM model's dynamic nature better accommodates the complex, iterative process of SRL in writing compared to traditional instruction or existing computer-based scaffolds (Lim et al., 2021; Zheng et al., 2023). This aligns with recent recommendations for optimizing SRL scaffolding that personalized scaffolds should provide support based on learners' actual SRL processes, not merely their learning progress, and fading of support as learners become capable of independent SRL (Lim et al., 2023; Van der Graaf et al., 2023). This comprehensive support likely facilitated the synergistic operation of key SRL strategies (Flower & Hayes, 1981).

6.2. Dynamics of writing motivation

This study revealed intriguing dynamics in students' writing motivation. Viewed through the lens of SCT, reciprocal determinism emphasizes the interplay among environmental, personal, and behavioral influences (Bandura, 1997). The interaction and timely feedback provided by the LLM positively impacted students' writing self-efficacy, aligning with Bandura's emphasis on mastery experiences as sources of self-efficacy (Liu et al., 2022). This resonates with previous findings (Bai et al., 2022) which suggested that providing immediate feedback on students' writing performance can positively affect students' adjustment of their beliefs about self-efficacy.

While self-efficacy was bolstered, a rapid waning in writing interest was observed. These results suggest that while the LLM approach has a direct and sustained impact on writing self-efficacy, the initial enhancement in writing interest it cultivates might be transient. Self-Determination Theory differentiates between intrinsic and extrinsic motivation (Deci & Ryan, 1985), and this observed outcome may align with the inherent nature of these motivational structures. Writing self-efficacy, reflecting beliefs about one's capabilities, manifests greater stability over time, especially in the LLM group. In contrast, writing interest (a more unstable intrinsic motivator) seems to wane over time, indicating inherent challenges in sustaining intrinsic motivation and underscoring the complexity of cultivating and maintaining writing motivation over the long haul (Boo et al., 2015; Renninger & Hidi, 2022). Future research could explore how LLMs can integrate features to sustain intrinsic motivation, perhaps through optimal motivation profiles or gamification elements (Huang et al., 2023).

6.3. The "humans in the loop" perspective on AI-assisted language instruction

This study highlights the potential of integrating LLM-based tutoring into language strategy instruction while emphasizing the crucial role of human educators. During the intervention, the authors observed that students sometimes experienced frustration or confusion with GPT-4's feedback or tasks, particularly when they perceived them as unclear or not directly applicable. This observation was confirmed by the CALLA-LLM instructor. In such situations, teachers needed to intervene promptly, offering both cognitive and emotional support, clarifying feedback, and adjusting tasks to maintain learner engagement.

The "Humans in the Loop" approach (Cardona et al., 2023), central to the pedagogical implications of this study, emphasizes the imperative for human intervention, oversight, and interaction in AI-driven instruction. In the CALLA-LLM model, teachers and LLMs play distinct but complementary roles. LLMs provide personalized scaffolding, real-time feedback, and assessment, automating many of the instructional tasks that would typically fall to the teacher in traditional CALLA models. This frees teachers to focus more on align AI-assisted instruction with overarching educational objectives, ensure coherence and continuity of the learning experience, and impart the humanistic care essential for fostering motivation and engagement. In contrast, in traditional CALLA models without AI support, teachers must shoulder the bulk of instructional scaffolding tasks, potentially diverting their attention from higher-order educational goals. The CALLA-LLM model thus optimizes the teacher's role through human-AI collaboration.

This teacher intervention is crucial from a cognitive-emotional dialectical perspective, as human educators help students navigate the cognitive challenges of AI-assisted learning while providing the emotional support and socio-cultural context necessary for deep understanding (Selwyn, 2019; Vygotskiĭ & Cole, 1978). Rather than viewing AI as a binary choice between machine efficiency and human expertise, the CALLA-LLM model presents a symbiotic enhancement that harnessing the strengths of both AI and human educators to optimize language learning outcomes.

7. Limitations and future research

This study has some limitations. First, the participants were drawn from a single primary school with relatively high English proficiency levels, which may limit the generalizability of the findings. Second, due to resource constraints, this study relied solely on quantitative data from questionnaires and did not collect qualitative data on the specific roles of teacher interventions and LLMs in the classroom.

Future research could adopting a mixed-methods design to gain a more comprehensive understanding of the impact of LLM-based instruction on students' learning processes. Additionally, Researchers could adopt a cognitive-emotional dialectical perspective to investigate how human educators and AI can work together to address learners' cognitive and emotional needs, fostering motivation, engagement, and deep understanding.

Disclosure statement

No competing financial interests exist.

Ethics approval

The experiment was approved by the Ethical Committee of Wenzhou University (WZU-2023-081) and was run in accordance with the Declaration of Helsinki.

Funding

This research was supported by the National Social Science Fund of China (BCA190089).

ORCID

Ze-Min Liu http://orcid.org/0000-0001-9331-1374
Xin-Dong Ye http://orcid.org/0000-0003-4187-0160

Data availability statement

The data presented in this study are available on request from the corresponding author.

References

An, X., Chai, C. S., Li, Y., Zhou, Y., & Yang, B. (2023). Modeling students' perceptions of artificial intelligence assisted language learning. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2023.2246519

Anthropic (2023). Claude 2. https://www.anthropic.com/news/claude-2

Bai, B. (2015). The effects of strategy-based writing instruction in Singapore primary schools. *System*, *53*, 96–106. https://doi.org/10.1016/j.system.2015.05.009

Bai, B., Guo, W., & Wang, C. (2024). Relationships between struggling EFL writers' motivation, self-regulated learning (SRL), and writing competence in Hong Kong primary schools. *Applied Linguistics Review, 15*(1), 135–159. https://doi.org/10.1515/applirev-2020-0131

Bai, B., & Wang, J. (2021). Hong Kong secondary students' self-regulated learning strategy use and English writing: Influences of motivational beliefs. *System*, *96*, 102404. https://doi.org/10.1016/j.system.2020.102404

Bai, B., Wang, J., & Zhou, H. (2022). An intervention study to improve primary school students' self-regulated strategy use in English writing through e-learning in Hong Kong. *Computer Assisted Language Learning*, *35*(9), 2265–2290. https://doi.org/10.1080/09588221.2020.1871030

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W H Freeman/Times Books/Henry Holt & Co.

Boo, Z., Dörnyei, Z., & Ryan, S. (2015). L2 motivation research 2005–2014: Understanding a publication surge and a changing landscape. *System*, *55*, 145–157. https://doi.org/10.1016/j.system.2015.10.006

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). *Sparks of Artificial General Intelligence: Early experiments with GPT-4* (arXiv:2303.12712). arXiv. https://doi.org/10.48550/arXiv.2303.12712

Cardona, M. A., Rodríguez, R. J., Ishmael, K. (2023). Artificial *intelligence and the future of teaching and learning*. U.S. Department of Education Office. https://tech.ed.gov/ai-future-of-teaching-and-learning/

Chamot, A. U. (2005). Language learning strategy instruction: Current issues and research. *Annual Review of Applied Linguistics*, *25*, 112–130. https://doi.org/10.1017/S0267190505000061

Chamot, A. U., & O'Malley, J. M. (1994). *The CALLA handbook: Implementing the cognitive academic language learning approach*. Addison-Wesley Pub. Co.

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. *Computers & Education*, *210*, 104967. https://doi.org/10.1016/j.compedu.2023.104967

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality*, *19*(2), 109–134. https://doi.org/10.1016/0092-6566(85)90023-6

Ebadijalal, M., & Moradkhani, S. (2023). Impacts of computer-assisted collaborative writing, collaborative prewriting, and individual writing on EFL learners' performance and motivation. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2023.2178463

Fathi, J., & Rahimi, M. (2022). Electronic writing portfolio in a collaborative writing environment: Its impact on EFL students' writing performance. *Computer Assisted Language Learning*. Advance online publication. https://doi.org/10.1080/09588221.2022.2097697

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Finlayson, K., & McCrudden, M. T. (2020). Teacher-implemented writing instruction for elementary students: A literature review. *Reading & Writing Quarterly*, *36*(1), 1–18. https://doi.org/10.1080/10573569.2019.1604278

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*(4), 365. https://doi.org/10.2307/356600

Graham, S., & Harris, K. R. (2003). Students with learning disabilities and the process of writing: A meta-analysis of SRSD studies. In *Handbook of learning disabilities* (pp. 323–344). The Guilford Press.

Graham, S., & Sandmel, K. (2011). The process writing approach: A meta-analysis. *The Journal of Educational Research*, *104*(6), 396–407. https://doi.org/10.1080/00220671.2010.488703

Granado-Peinado, M., Cuevas, I., Olmos, R., Martín, E., Casado-Ledesma, L., & Mateos, M. (2023). Collaborative writing of argumentative syntheses by low-performing undergraduate writers: Explicit instruction and practice. *Reading and Writing*, *36*(4), 909–936. https://doi.org/10.1007/s11145-022-10318-x

Hahahumble (2023). SpeechGPT. GitHub. Retrieved from https://github.com/hahahumble/speechgpt

Harris, K. R., Santangelo, T., & Graham, S. (2010). Metacognition and strategies instruction in writing. In *Metacognition, strategy use, and instruction* (pp. 226–256) The Guilford Press.

Huang, A. Y. Q., Lu, O. H. T., & Yang, S. J. H. (2023). Effects of artificial intelligence–enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education*, *194*, 104684. https://doi.org/10.1016/j.compedu.2022.104684

Hyland, K. (2003). *Second language writing*. Cambridge University Press.

Ilieva, G., Yankova, T., Klisarova-Belcheva, S., Dimitrov, A., Bratkov, M., & Angelov, D. (2023). Effects of generative chatbots in higher education. *Information*, *14*(9), 492. https://doi.org/10.3390/info14090492

Jiang, W., & Eslami, Z. R. (2022). Effects of computer-mediated collaborative writing on individual EFL writing performance. *Computer Assisted Language Learning*, *35*(9), 2701–2730. https://doi.org/10.1080/09588221.2021.1893753

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, *54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Lim, L., Bannert, M., Van Der Graaf, J., Singh, S., Fan, Y., Surendrannair, S., Rakovic, M., Molenaar, I., Moore, J., & Gašević, D. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. *Computers in Human Behavior*, *139*, 107547. https://doi.org/10.1016/j.chb.2022.107547

Lim, L.-A., Gentili, S., Pardo, A., Kovanović, V., Whitelock-Wainwright, A., Gašević, D., & Dawson, S. (2021). What changes, and for whom? A study of the impact of learning analytics-based process feedback in a large course. *Learning and Instruction*, *72*, 101202. https://doi.org/10.1016/j.learninstruc.2019.04.003

Lin, M. P.-C., & Chang, D. (2023). CHAT-ACTS: A pedagogical framework for personalized chatbot to enhance active learning and self-regulated learning. *Computers and Education: Artificial Intelligence*, *5*, 100167. https://doi.org/10.1016/j.caeai.2023.100167

Liu, Z.-M., & Chen, Y.-H. (2022). A modularity design approach to behavioral research with immersive virtual reality: A SkyrimVR-based behavioral experimental framework. *Behavior Research Methods, 55*, 3805–3819. https://doi.org/10.3758/s13428-022-01990-6

Liu, Z.-M., Fan, X., Liu, Y., & Ye, X. (2022). Effects of immersive virtual reality cardiopulmonary resuscitation training on prospective kindergarten teachers' learning achievements, attitudes and self-efficacy. *British Journal of Educational Technology*, *53*(6), 2050–2070. https://doi.org/10.1111/bjet.13237

MacArthur, C. A., Philippakos, Z. A., & Graham, S. (2016). A multicomponent measure of writing motivation with basic college writers. *Learning Disability Quarterly*, *39*(1), 31–43. https://doi.org/10.1177/0731948715583115

OpenAI. (2023). *GPT-4 Technical Report* (arXiv:2303.08774). arXiv. https://doi.org/10.48550/arXiv.2303.08774

Ou, W. A., Gu, M. M., & Hult, F. M. (2023). Translanguaging for intercultural communication in international higher education: Transcending English as a lingua franca. *International Journal of Multilingualism*, *20*(2), 576–594. https://doi.org/10.1080/14790718.2020.1856113

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40. https://doi.org/10.1037/0022-0663.82.1.33

Rahimi, M., & Fathi, J. (2022). Exploring the impact of wiki-mediated collaborative writing on EFL students' writing performance, writing self-regulation, and writing self-efficacy: A mixed methods study. *Computer Assisted Language Learning*, *35*(9), 2627–2674. https://doi.org/10.1080/09588221.2021.1888753

Renninger, K. A., & Hidi, S. E. (2022). Interest: A unique affective and cognitive motivational variable that develops. In *Advances in motivation science* (Vol. 9, pp. 179–239). Elsevier. https://doi.org/10.1016/bs.adms.2021.12.004

Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education* (1st ed.). Polity Press.

Silitonga, L. M., Hawanti, S., Aziez, F., Furqon, M., Zain, D. S. M., Anjarani, S., & Wu, T.-T. (2023). The impact of AI chatbot-based learning on students' motivation in english writing classroom. In Y.-M. Huang, & T. Rocha (Eds.), *Innovative technologies and learning* (vol. 14099, pp. 542–549). Springer. https://doi.org/10.1007/978-3-031-40113-8_53

Stojanov, A. (2023). Learning with ChatGPT 3.5 as a more knowledgeable other: An autoethnographic study. *International Journal of Educational Technology in Higher Education*, *20*(1), 35. https://doi.org/10.1186/s41239-023-00404-7

Sun, T., & Wang, C. (2020). College students' writing self-efficacy and writing self-regulated learning strategies in learning English as a foreign language. *System*, *90*, 102221. https://doi.org/10.1016/j.system.2020.102221

Sun, T., Wang, C., & Wang, Y. (2022). The effectiveness of self-regulated strategy development on improving English writing: Evidence from the last decade. *Reading and Writing*, *35*(10), 2497–2522. https://doi.org/10.1007/s11145-022-10297-z

Teng, L. S. (2022). Explicit strategy-based instruction in L2 writing contexts: A perspective of self-regulated learning and formative assessment. *Assessing Writing*, *53*, 100645. https://doi.org/10.1016/j.asw.2022.100645

Teng, L. S., & Zhang, L. J. (2020). Empowering learners in the second/foreign language classroom: Can self-regulated learning strategies-based writing instruction make a difference? *Journal of Second Language Writing*, *48*, 100701. https://doi.org/10.1016/j.jslw.2019.100701

The jamovi Project (2022). *Jamovi* (2.3) [Computer software]. https://www.jamovi.org/

Troia, G. A., Harbaugh, A. G., Shankland, R. K., Wolbers, K. A., & Lawrence, A. M. (2013). Relationships between writing motivation, writing activity, and writing performance: Effects of grade, sex, and ability. *Reading and Writing*, *26*(1), 17–44. https://doi.org/10.1007/s11145-012-9379-2

Van Der Graaf, J., Raković, M., Fan, Y., Lim, L., Singh, S., Bannert, M., Gašević, D., & Molenaar, I. (2023). How to design and evaluate personalized scaffolds for self-regulated learning. *Metacognition and Learning*, *18*(3), 783–810. https://doi.org/10.1007/s11409-023-09361-y

Vygotskiĭ, L. S., & Cole, M. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Zhang, R., & Zou, D. (2022). Self-regulated second language learning: A review of types and benefits of strategies, modes of teacher support, and pedagogical implications. *Computer Assisted Language Learning*, *37*(4), 720–765. https://doi.org/10.1080/09588221.2022.2055081

Zheng, J., Lajoie, S. P., Wang, T., & Li, S. (2023). Supporting self-regulated learning in clinical problem-solving with a computer-based learning environment: The effectiveness of scaffolds. *Metacognition and Learning*, *18*(3), 693–709. https://doi.org/10.1007/s11409-023-09352-z

Zimmerman, B. J. (2000). Attaining self-regulation. In *Handbook of self-regulation* (pp. 13–39). Elsevier. https://doi.org/10.1016/B978-012109890-2/50031-7

Zimmerman, B. J., & Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. *American Educational Research Journal*, *31*(4), 845–862. https://doi.org/10.3102/00028312031004845