

Exploring the Impact of Integrated AWE and Generative AI Feedback on Chinese EFL Undergraduates' Higher-Order Thinking in Argumentative Writing

SAGE Open
January-March 2026: 1–12
© The Author(s) 2026
DOI: 10.1177/21582440251413884
journals.sagepub.com/home/sgo


Hongxia Hao¹ , Abu Bakar Razali¹ , and Ruijia Zuo² 

Abstract

Although automated writing evaluation (AWE) and artificial intelligence (AI) tools have been widely practiced in EFL/ESL writing instruction, there is a lack of empirical research on the effect of the integration of both AWE and AI feedback on students' higher-order thinking (HOT). Therefore, this study is to explore the impact of integrating AWE and AI feedback on Chinese EFL undergraduates' higher-order thinking (HOT) in argumentative writing based on Revised Bloom's Taxonomy and Cognitive Feedback Theory. Pre- and post-tests and semi-structured interviews were used to study 64 third-year students in the English major at a Chinese public university for 16 weeks. The experimental group ($n = 32$) received AWE (Pigai) and AI (ChatGPT) feedback, while the control group ($n = 32$) received only AWE (Pigai) feedback. Quantitative results showed that EG students had significant improvements in higher-order thinking (HOT; analysis, evaluation, and creation; $p < .001$) with a high effect size ($d > 0.80$), while the CG students had a smaller improvement ($d > 0.15$). ANOVA confirmed that analysis had the highest effect size ($p < .001$, $\eta^2 = .862$), followed by evaluation ($p < .001$, $\eta^2 = .818$) and creation ($p < .001$, $\eta^2 = .812$). Qualitative results showed that AWE and AI tools were complementary, in which AWE could help students correct superficial language errors, but AI could improve students' higher-order thinking (HOT) in analysis, evaluation, and creation. They can focus on language and higher-order thinking (HOT) and optimized revision strategies. However, students also faced problems in understanding feedback and over-reliance on it.

Keywords

artificial intelligence (AI), automated writing evaluation (AWE), undergraduates, higher-order thinking, argumentative writing

Introduction

Technology and education are constantly integrating and developing, especially in teaching English as a second foreign language/teaching English as a foreign language (ESL/EFL; Solikhah, 2023). Automated writing evaluation (AWE) tools and generative artificial intelligence (AI) tools are widely used in EFL/ESL writing instruction. AWE tools provide learners with immediate formative feedback on grammar, vocabulary, structure, and other aspects (e.g., Criterion, Pigai, Grammarly), and are widely used in large-class EFL/ESL writing instruction in higher education (Nunes et al., 2022; Ranalli, 2021). In contrast, generative artificial intelligence (AI) tools (such as ChatGPT) can simulate dialogue and interact with students through iterative feedback, helping to cultivate their critical thinking, creativity, and problem-solving skills.

They have been incorporated into writing instruction in higher education in many countries (Borge et al., 2024; Naznin et al., 2025). Therefore, AWE focuses on language accuracy, while AI supports creativity. Integrating AWE and AI tools may provide more comprehensive feedback for ESL/EFL learners' writing.

¹Universiti Putra Malaysia, Serdang, Selangor, Malaysia

²Xihua University, Chengdu, Si Chuan, China

Corresponding Author:

Hongxia Hao, Department of Language and Humanities Education, Faculty of Educational Studies, Universiti Putra Malaysia, Serdang, Selangor 43400, Malaysia.

Email: haohongxia.tesl@gmail.com

Data Availability Statement included at the end of the article



The integration of the technology is crucial for developing learners' argumentative writing skills, as it requires both language accuracy and higher-order thinking (HOT) such as analysis, evaluation, and creation (Mokhtar et al., 2020). However, research indicates that Chinese EFL university faculty often judge undergraduates' writing based on language accuracy, while undergraduates demonstrate weaknesses in logical reasoning, argumentation, and critical thinking in argumentative writing (Y. Li et al., 2022). This results in insufficient creativity and critical thinking in students' argumentative writing.

Higher-order thinking (HOT) is not only a process by which learners analyze, evaluate, synthesize, create, and solve problems, but also an important indicator of their critical thinking, complex problem-solving, and innovation abilities (Anderson & Krathwohl, 2001). However, higher-order thinking (HOT) is often neglected in argumentative writing among Chinese English as a Foreign Language (EFL) students, resulting in weaker argumentation, refutation, and creativity (Pei et al., 2017). Chinese EFL undergraduates lack supporting arguments and reverse thinking in their essays, leading to insufficient evaluation, reconstruction, and differentiation of existing viewpoints (Wu et al., 2023). Although some universities have applied AWE tools to argumentative writing instruction, such as Pigai (Liu et al., 2022), the effect of AWE tools alone on cultivating students' higher-order thinking (HOT) is limited (M. Li et al., 2025). Therefore, given the inadequacy of higher-order thinking (HOT) cultivation in Chinese EFL writing instruction, the authors of this study believe that combining AWE with artificial intelligence (AI) feedback may better guide students to develop higher-order thinking (HOT).

Technology-assisted and EFL/ESL writing feedback is popular, but previous studies have separately explored the advantages of AWE tools in language feedback (Fan, 2023; Fu & Liu, 2022) and the potential of AI in generating content and logical structure (Mahapatra, 2024; Suh et al., 2025). The impact of AWE tools on EFL/ESL students' writing is mainly focused on language-level feedback (i.e., spelling, grammar, and syntax), but there is a lack of research on higher-order thinking (HOT; Analyze–Evaluate–Create), such as critical logic and refutation ability (Wambsganss et al., 2022). Meanwhile, generative AI tools are currently mostly used in pre-writing conception and brainstorming and have limited intervention in the writing process feedback (Suh et al., 2025). In addition, most existing studies study AWE and AI tools separately, lacking empirical exploration of integrating the feedback advantages of both (Shi & Aryadoust, 2024). Few previous studies have used a mixed methods design approach to explore the

impact of the integration of AWE and AI on Chinese EFL undergraduates' higher-order thinking (HOT) in argumentative writing. Chinese EFL undergraduates lack logical and critical thinking in argumentative writing, and how to improve this problem requires more research to be done.

Therefore, this study used the integration of AWE and AI feedback to explore their impact on students' higher-order thinking (HOT) in English academic argumentative writing. The two research questions of this study are as follows:

1. What is the impact of integrated AWE and generative AI feedback on Chinese EFL undergraduates' higher-order thinking in English argumentative writing?
2. How do Chinese EFL undergraduates perceive the impacts of integrated AWE and AI feedback in supporting their higher-order thinking?

Literature Review

Argumentative Writing in EFL/ESL and Higher-Order Thinking

Argumentative writing is a genre of EFL/ESL English academic writing that requires learners to have clear arguments, evidence, and argumentation structures (Lee & Lee, 2024). Argumentative writing is essentially a process of cultivating students' higher-order thinking (HOT), namely analysis, evaluation, and creation. In the revised Bloom's Taxonomy of Cognitive Objectives, higher-order thinking (HOT) refers to analysis, evaluation, and creation (Anderson & Krathwohl, 2001). In this study, analysis refers to the process by which students clearly and logically analyze complex arguments, evidence, and argumentative structures. Evaluation refers to the process by which students critically evaluate the validity and rationality of arguments and propose reasonable rebuttals. Creativity refers to the process by which students integrate multiple perspectives to form an original, independent, and novel argumentative structure.

In English as a Foreign/Second Language (EFL/ESL) argumentative writing, higher-order thinking (HOT) is closely linked to critical thinking, logical reasoning, and persuasiveness. However, Chinese EFL undergraduates face numerous challenges in argumentative writing, such as insufficient language knowledge, inadequate argumentation skills, unclear logical structure, and weak critical thinking abilities (Peng & Bao, 2023). Therefore, targeted strategies are needed to address the problems faced by Chinese EFL undergraduates in argumentative writing, namely, focusing on cultivating students' higher-order thinking (HOT). This study combines AWE (Analytical,

Evaluative, and Artificial) feedback and AI feedback, which may play a positive role in cultivating students' higher-order thinking (HOT), namely their analytical, evaluative, and constructive abilities.

AWE and Generative AI Tools in EFL/ESL Argumentative Writing

The integration of technology with English as a Foreign Language/Second Language (EFL/ESL) writing instruction is becoming increasingly significant, particularly with AWE and AI. AWE tools (such as Pigai and Grammarly) provide immediate and objective feedback on students' grammar, spelling, vocabulary, and syntax in their writing (Fan, 2023). This has a positive effect on reducing teachers' workload and improving students' learning efficiency (Fu & Liu, 2022). However, AWE tools focus on low-level language feedback, offering limited feedback on the logic, coherence, and plausibility of arguments, points, and evidence in students' argumentative writing (Wambsganss et al., 2022; Yang et al., 2024). Therefore, using a single AWE tool may have a limited effect on improving students' higher-order thinking (HOT), such as analysis, evaluation, and creativity, in argumentative writing.

Generative AI tools (such as ChatGPT) are significant for improving students' higher-order thinking (HOT) in argumentative writing, including argumentation, evidence gathering, and rebuttal (Suh et al., 2025). Generative AI enhances argumentation, critical thinking, and reflective abilities by simulating human interaction, thus strengthening creativity and positively impacting higher-order thinking (HOT; Janse van Rensburg, 2024; Mahapatra, 2024). However, relying solely on AI feedback may lead to over-reliance on it and a lack of higher-order thinking (HOT; Chan & Hu, 2023; Kim et al., 2024).

In conclusion, AWE tools and AI tools are highly complementary, and combining them can fully leverage their respective strengths. Current research studies AWE and AI feedback separately, lacking empirical studies on the impact of their combination on higher-order thinking (HOT) in argumentative writing among undergraduate English as a Foreign Language/Second Language (EFL/ESL) students. Therefore, this study aims to bridge the gap between theory and practice by integrating AWE and generative AI feedback in order to support Chinese EFL undergraduates in higher-order thinking (HOT) in argumentation writing.

Feedback and Higher-Order Thinking in Argumentative Writing

Feedback plays a crucial role in cultivating students' higher-order thinking (HOT) during the argumentative

writing revision process. Feedback provides learners with opportunities to review writing goals, monitor progress, and revise (Hyland & Hyland, 2006). Superficial feedback focuses on the linguistic level (grammar, vocabulary, syntax), while deep feedback focuses on arguments, evidence, and argumentation structure (Hyland & Hyland, 2006). Effective feedback is characterized by clarity, actionability, task-specific focus, and in-depth processing to stimulate students' reflective, evaluative, and reconstructive thinking (Shute, 2008). In this study, the AWE tool addresses surface-level error correction, while the generative AI tool addresses deep-level error correction. This complementary feedback mechanism aligns with higher-order thinking (HOT), potentially maximizing its effectiveness.

Research Gap

Technology-assisted feedback for EFL/ESL writing is becoming increasingly common, but previous research has primarily focused on the language-level feedback (spelling, grammar, and syntax) provided by AWE tools (Fan, 2023; Fu & Liu, 2022), while research on higher-order thinking (HOT; analysis-evaluation-creation) remains insufficient (Wambsganss et al., 2022). Furthermore, generative AI tools currently focus mainly on feedback at the content and logical structure levels (Mahapatra, 2024; Suh et al., 2025), and are mostly applied to pre-writing brainstorming and conception, with limited intervention in feedback during the writing process (especially the revision stage; Suh et al., 2025). Moreover, existing research largely examines AWE and AI tools separately, lacking empirical exploration of integrating the advantages of both feedback methods (Shi & Aryadoust, 2024). Few studies have employed a mixed-method design to explore the impact of the integration of AWE and AI on higher-order thinking (HOT) in EFL argumentative writing among Chinese undergraduate students. Therefore, this study takes the integration of AWE and AI as an example to explore its impact on higher-order thinking (HOT) in argumentative writing among Chinese EFL undergraduates.

Theoretical Framework

The Revised Bloom's Taxonomy (Anderson & Krathwohl, 2001) and Cognitive Feedback Theory (Shute, 2008) were used as theoretical foundations to explain how the integration of AWE and generative AI feedback can support Chinese EFL undergraduates' higher-order thinking (HOT) in argumentative writing. The Revised Bloom's Taxonomy was revised by Anderson and Krathwohl in 2001, and it mainly included six progressive cognitive dimensions: remember,

understand, apply, analyze, evaluate, and create. Among them, analyze, evaluate, and create are high-order thinking skills. Argumentative writing under the Revised Bloom's Taxonomy is a process of analysis-evaluation-creation. In argumentative writing, students analyze the arguments, evidence, and logical relationships (analyze), then critically evaluate the validity, originality, and rationality of their argument structure (evaluate), and finally construct their own argument structure with originality, uniqueness, and diversity (create). However, Chinese EFL undergraduates, under the test-oriented education environment, focus too much on language accuracy rather than higher-order thinking. This highlights the importance and necessity of teaching interventions to cultivate students' higher-order thinking. Therefore, the Revised Bloom's Taxonomy provides a powerful cognitive framework for this study to analyze the high-order thinking process of analysis, evaluation, and creation of students in argumentative writing. This study used different types of feedback (AWE-only vs. AWE + AI) to explore its impact on students' high-order thinking (HOT) in argumentative writing.

Meanwhile, Cognitive Feedback Theory is introduced to construct a cognitive feedback framework. It emphasizes formative feedback and has the significant characteristics of timelines, actionability and cognitive engagement (Shute, 2008). Under this theory, learners can modify both the surface features and the deep structure of a text. According to the specific and clear guidance of the feedback, they can directly modify the task, and finally promote learners' thinking processing under the feedback. AWE feedback helps learners reduce surface-level errors (grammar, vocabulary, and structure), while generative AI feedback provides learners with higher-order thinking (HOT). The integration of AWE and AI feedback forms a complementary cycle of surface correction, cognitive activation, and deeper reasoning. Therefore, cognitive feedback theory provides a powerful cognitive framework for this study to analyze students' higher-order thinking (HOT) processes in argumentative writing. This study used different types of feedback (AWE-only vs. AWE + AI) to explore its impact on students' higher-order thinking (HOT) in argumentative writing. Based on the integration of Revised Bloom's Taxonomy and Cognitive Feedback Theory, researchers proposed a process of feedback, cognitive activation, and improvement of higher-order thinking (HOT).

These two theories are uniquely tested in the Chinese EFL context. Chinese undergraduate writing instruction overemphasizes linguistic accuracy, while lacking attention to higher-order thinking (HOT) such as argumentation, refutation, and originality (Wu et al., 2023). Chinese EFL undergraduates lack higher-order thinking

(HOT) in argumentative writing in large classes. In this context, integrating AWE with generative AI feedback facilitates personalized, diverse, and high-level feedback. Therefore, Chinese EFL undergraduates are suitable subjects for this study to explore how integrated feedback promotes the development of higher-order thinking (HOT). The integration of AWE and generative AI feedback—AWE feedback (surface scaffolding) reduces language barriers and cognitive load, while generative AI feedback (deep stimulation) promotes argument analysis, critical evaluation, and creative synthesis—provides a solid theoretical foundation for exploring how different forms of feedback interact to promote higher-order thinking (HOT) in argumentative writing among Chinese EFL undergraduates.

Methodology

Research Design

The researchers adopted an explanatory sequential mixed method design to explore the impact of integrating AWE and generative AI feedback on Chinese EFL undergraduates' higher-order thinking (HOT) in argumentative writing. In the quantitative stage, the researchers used a pre-post-test to collect the impact of different feedback types on students' HOT in argumentative writing. Students in the experimental group (EG) and the control group (CG) were required to complete the pre-test and post-test respectively. EG received AWE and AI feedback during the intervention period, while CG only received AWE feedback. In addition, in the qualitative stage, the researchers conducted semi-structured interviews with EG students to explore their views on the impact of integrated feedback on their HOT.

Participants

Sixty-four third-year undergraduates majoring in English at a public university in China were selected as subjects for this study. All participants had received 2 years of English major study and had a solid writing foundation (all passed the College English Test Band 6). Students were randomly divided into an experimental group (EG, $n = 32$) and a control group (CG, $n = 32$). EG students received AWE and generative AI feedback (Pigai + ChatGPT) during the 16-week intervention period. CG students only received AWE feedback (Pigai) during the 16-week intervention period.

Instrument and Procedure

The two writing tasks in IELTS Writing Task 2 were used for both the pre-test and post-test. The main reason for using the argumentative writing section of IELTS

Writing Task 2 as a pre-test and post-test is that it requires students to analyze, evaluate, and create in their writing. This cognitive process is highly relevant to higher-order thinking (HOT) and is characterized by standardization and international recognition. Students need to identify and analyze the prompt, distinguish positions, organize arguments (analysis), evaluate viewpoints, logical reasoning, and strengthen the evidence (evaluation), synthesize multiple viewpoints, and propose an independent, persuasive, and innovative position (creativity). The “task response” and “coherence and cohesion” elements in the IELTS writing assessment criteria require students to use logical reasoning, critically understand the prompt, and support their viewpoints with evidence and examples, reflecting evaluative and analytical thinking. Coherence and cohesion emphasize the organization and integration of ideas, which aligns with the comprehensive dimensions of higher-order thinking (HOT). Therefore, IELTS Writing Task 2 can assess both students’ language proficiency and their level of higher-order thinking (HOT) development in argumentative writing.

In both the pre-test and post-test, students were required to complete a 250 to 300-word argumentative essay within 40 min. To ensure that the differences were attributable to the intervention rather than external factors, all participating classes took the tests in a pre-designated classroom with identical instructions, time limits, tests, and invigilation procedures. Both tests consisted of argumentative essays of comparable difficulty. Students’ writing scores in both tests were graded according to the IELTS scoring criteria. The Higher-order thinking (HOT) assessment used Bloom’s Taxonomy of Cognitive Objectives (Anderson & Krathwohl, 2001). Pre-test scores served as a benchmark for language and cognitive abilities in selected classes; follow-up studies were conducted if classes were deemed comparable. Post-test scores indicated whether HOT scores showed improvement over pre-test scores. To ensure objectivity, two teachers with over 10 years of experience teaching argumentative writing served as graders, using the same scoring criteria to grade all students’ HOT scores twice. Each dimension was scored on a 5-point scale (1 being the lowest and 5 the highest). Any discrepancy exceeding one level was resolved through discussion and negotiation between the two graders. Both the scoring training and pilot scoring were conducted by two scorers (Cohen’s $\kappa > 0.80$), which demonstrates a high degree of consistency among the raters.

The researchers standardized the experimental procedure, using ChatGPT (version 4.0) as the AI tool. Students in the experimental group received feedback from ChatGPT by inputting the same prompts (e.g.,

identifying and analyzing positions, organizing arguments (analysis), evaluating viewpoints, logical reasoning, and strengthening evidence (evaluation), synthesizing multiple viewpoints, and proposing an independent, persuasive, and innovative position (creativity)). To ensure the quality of feedback, all ChatGPT outputs were ultimately reviewed by the instructors. Experimental group (EG) students submitted their initial drafts to the AWE (Pigai) tool, which provided automated feedback (e.g., grammar, vocabulary, and structure). Students then revised their essays based on the AWE feedback and submitted the revised text to the AI tool (ChatGPT) according to specified instructions. The AI tool (ChatGPT) provided in-depth feedback based on the students’ instructions. Finally, the teachers conducted a final manual review of the AI’s output, and the students then revised and refined their work. Students in the control group (CG) received only feedback from the AWE (Pigai) tool, which primarily focused on surface in their essays. All students write under the same conditions each week (including time, language, word count, materials, and testing environment).

Data Collection

The argumentative essay section of IELTS Writing Task 2 was chosen for both the pre-test and post-test primarily because it requires students to utilize analytical, evaluative, and creative abilities in their writing. This cognitive process is highly correlated with higher-order thinking (HOT) and is standardized and internationally recognized. In both the pre-test and post-test, students were required to complete a 250 to 300-word argumentative essay within 40 min. To ensure that differences were due to intervention rather than external factors, all participating classes were tested in pre-designated classrooms with identical testing requirements, time limits, test papers, and invigilation procedures. HOT assessment used Bloom’s Taxonomy of Cognitive Objectives (Anderson & Krathwohl, 2001). To ensure objectivity, two teachers with over 10 years of experience teaching argumentative writing served as graders, using the same scoring criteria to score all students’ HOT scores twice. Each dimension was scored on a 5-point scale (1 being the lowest and 5 the highest). Any discrepancies exceeding one level were resolved through discussion and consultation between the two graders. The scoring training, and trial assessments were all conducted by two scorers (Cohen’s $\kappa > 0.80$), indicating a high degree of consistency among the scorers.

Nine students from the experimental group (EG)-three each of high, medium, and low writing scores were randomly selected for one-on-one semi-structured

Table 1. Descriptive Statistics.

Group	Test	M	SD	Test	M	SD	N
EG	Pre-analyze	2.81	1.06	Post-analyze	3.75	1.02	32
	Pre-evaluate	2.78	0.94	Post-evaluate	3.72	1.14	32
	Pre-create	2.84	0.92	Post-create	3.72	1.17	32
CG	Pre-analyze	2.81	1.06	Post-analyze	3.06	1.08	32
	Pre-evaluate	2.75	1.02	Post-evaluate	2.91	1.09	32
	Pre-create	2.78	1.04	Post-create	3.06	1.01	32

interviews. Each interview lasted approximately 20 to 30 min and was recorded and transcribed verbatim. The interview questions were original, totaling 12 questions, each corresponding to a specific higher-order thinking (HOT) dimension. There were three questions on each of the four themes: students' overall experience with the combination of AWE and AI feedback; the effect of combining AWE and AI feedback on activating higher-order thinking (HOT); the impact of AWE and AI feedback on changes in strategy modification and thought processes; and challenges and suggestions regarding AWE and AI feedback. The interview results were double-coded by two researchers. To identify students' relevant cognitions about higher-order thinking (HOT) and feedback, and to ensure the rigor and transparency of the methodology, theme analysis was used.

Data Analysis

Descriptive statistics, paired sample *t*-test, and analysis of covariance (ANCOVA) were used in quantitative analysis. Descriptive statistics and paired sample *t*-test were used to compare the HOT scores of each group in the pre-test and post-test. ANCOVA was used to compare post-test scores between groups and control for potential confounding variables.

Theme analysis and the six-step approach proposed by Braun and Clarke (2006; familiarization with data, initial coding, topic generation, topic review, topic definition, and report writing) were used to systematically identify, organize, and interpret students' perceptions of AWE and AI feedback in argument writing. Two coders performed coding separately to ensure the validity and reliability of the analysis results. Triangulation of quantitative and qualitative data helped validate the research findings. The researchers also combined topic analysis with deductive and inductive coding, i.e., inductive coding based on predefined dimensions of higher-order thinking (HOT; analysis, evaluation, creation) while allowing students to identify new patterns in their responses. This helps researchers accurately capture students' views on the four themes.

Ethical Consideration

All procedures in this study were conducted in accordance with the ethical standards of the target university. All participants were required to sign a written informed consent form before participation, which included the research objectives, procedures, potential risks and benefits, confidentiality measures, and the right to withdraw at any time. Participants' personal information was anonymized, and all data was securely stored in accordance with the institution's data protection guidelines. This study involved writing tests and semi-structured interviews with participants, posing minimal psychological or physiological risks. The benefits of this study outweigh the risks to participants; it may improve their writing skills and enhance higher-order thinking abilities, providing valuable insights for educational practice.

Results and Findings

Quantitative Results

Combining AWE and AI feedback improved higher-order thinking (HOT) in argumentative writing for the experimental group (EG) students (Table 1). The most significant improvement was in analysis, with the average score increasing from 2.81 ($SD = 1.06$) to 3.75 ($SD = 1.02$), indicating that students were able to identify viewpoints, differentiate positions, and organize arguments, demonstrating enhanced analysis. In addition, evaluation improved, with the average score increasing from 2.78 ($SD = 0.94$) to 3.72 ($SD = 1.14$), indicating enhanced evaluation in assessing viewpoints, logical reasoning, and strengthening evidence. Finally, creation improved, with the average score increasing from 2.84 ($SD = 0.92$) to 3.72 ($SD = 1.17$), demonstrating improved creativity in synthesizing multiple viewpoints and proposing independent, persuasive, and innovative positions. While the control group (CG) students showed a slight improvement in higher-order thinking (HOT; analysis, evaluation, and creativity) in argumentative writing, this improvement was significantly lower than that of the experimental group (EG),

Table 2. Paired Sample t-Test.

Group	Test pair	Dimension	<i>MD</i>	<i>t</i>	<i>df</i>	<i>p</i>	Cohen's <i>d</i>
EG	Pre vs. Post	Analyze	−0.94	−15.00	31	<.001	0.90
		Evaluate	−0.94	−10.52	31	<.001	0.90
		Create	−0.88	−10.06	31	<.001	0.84
CG	Pre vs. Post	Analyze	−0.25	−3.21	31	.002	0.23
		Evaluate	−0.16	−1.94	31	.029	0.15
		Create	−0.28	−3.48	31	<.001	0.27

Table 3. ANCOVA.

Dimension	Source	SS	<i>Df</i>	MS	<i>F</i>	<i>p</i>	Partial η^2
Analyze	Pre-test	58.50	1	58.50	380.40	<.001	.862
	Group	7.56	1	7.56	49.18	<.001	.446
	Error	9.38	61	0.154			
Evaluate	Pre-test	63.15	1	63.15	274.40	<.001	.818
	Group	9.74	1	9.74	42.40	<.001	.410
	Error	14.04	61	0.23			
Create	Pre-test	60.38	1	60.38	263.70	<.001	.812
	Group	5.63	1	5.63	24.58	<.001	.287
	Error	13.97	61	0.23			

indicating that AWE feedback alone has limited effectiveness in stimulating students' higher-order thinking (HOT). Furthermore, the experimental and control groups showed little difference in the pre-test, confirming their comparable starting points. However, in the post-test, the experimental group (EG) significantly outperformed the control group (CG) in higher-order thinking (HOT; analysis, evaluation, and creativity), demonstrating that combining AWE and AI feedback is more effective in improving students' higher-order thinking (analysis, evaluation, and creation) in argumentative writing.

Combining AWE and AI feedback improved the higher-order thinking (HOT; analysis, evaluation, and creation) of the experimental group (EG) students in argumentative writing ($p < .001$), with a large effect size; Cohen's d values were all greater than 0.80 ($d = 0.90/0.90/0.84$; Table 2). This suggests that the combination of AWE and AI feedback has a positive effect on enhancing students' deeper cognitive engagement. The experimental group (EG) students demonstrated stronger abilities to identify and analyze viewpoints, differentiate positions, organize arguments (analysis), evaluate viewpoints, reason logically, and strengthen evidence (evaluation), synthesize multiple viewpoints, and propose independent, persuasive, and innovative positions (creativity). In contrast, AWE feedback had a significant impact on the control group (CG) students ($p = .002/.029/.001$), but the effect size was small ($d = 0.23/0.15/0.27$), indicating that

its impact on students' higher-order thinking (HOT) was minimal. This further confirms that AWE feedback alone has little effect on students' higher-order thinking (HOT) in argumentative writing.

Pre-test scores are important covariates for higher-order thinking (HOT; analysis, evaluation, and creation; Table 3). Analytical ability is the strongest predictor of post-test scores ($p < .001$, $\eta^2 = .862$), followed by evaluation ability ($p < .001$, $\eta^2 = .818$), and lastly creativity ability ($p < .001$, $\eta^2 = .812$). This suggests that students with stronger initial analytical skills are more likely to benefit from subsequent feedback interventions. Furthermore, different feedback significantly affected higher-order thinking (HOT; analysis, evaluation, and creativity) in argumentative writing for both groups of students ($p < .001$), with effect sizes ranging from moderate to large ($\eta^2 = .446/.410/.287$). This confirms that the integration of AWE and AI feedback provides statistically significant improvements and has a profound impact on students' analytical, evaluative, and creative thinking in argumentative writing. This comprehensive feedback promotes students' higher-level cognitive development.

Qualitative Results

The researchers purposely selected nine students from EG for semi-structured interviews. There were three high-level, three medium-level, and three low-level students. The researchers used the theme analysis

framework of Braun and Clarke (2006) to derive four main themes from overall experience, higher-order thinking, thinking process, and challenge suggestions.

Theme 1: Overall Experience. Integration of AWE (Pigai) and AI (ChatGPT) provided different help and guidance for students' argumentative writing revisions. Students believed that AWE could improve their grammar, vocabulary, and sentence accuracy (surface modification). For example, a low-level student said, "AWE (Pigai) can immediately point out grammar, vocabulary, and other problems in my writing and give revision suggestions." However, AI feedback helps to improve students' higher-order thinking (HOT; analysis, evaluation, and creation) in argumentative writing and stimulates their deeper thinking. For example, a high-level student said: "ChatGPT suggested that I add rebuttals or stronger evidence. I will conduct in-depth analysis and thinking, and finally form creative arguments and structures." Most students believe that the integration of AWE and AI feedback can achieve the greatest effect. As one intermediate student said: "The integration of AWE (Pigai) and AI (ChatGPT) can improve my language ability and higher-order thinking ability." Therefore, these findings further confirmed that AWE focuses on surface language feedback and AI provides high-level deep feedback, and these two feedback methods have their own advantages and are complementary.

Theme 2: Activating Higher-Order Thinking Skills. The interview found that the integration of AWE and AI feedback can help students reconstruct argument structure, identify logical problems, and clarify argument points. As one intermediate student said, "Pigai points out my language problems, ChatGPT can point out that my argument structure and arguments are not clear. I think deeply about how to construct my writing logically based on its feedback." This confirmed that integration of AWE and AI feedback was conducive to improving students' ability to analyze structure and logical analysis. In addition, integration of AWE and AI feedback was conducive to improving the ability of students of different levels to question assumptions, evaluate evidence, and consider rebuttals. For example, a low-level student said, "After Pigai helped me clear the language barrier, ChatGPT would raise a hypothesis. What if someone disagrees? Is the evidence strong enough? I will think deeply about it and evaluate whether my argument needs to be supported by counterexamples." This confirmed that integration of AWE and AI feedback was conducive to improving students' deeper evaluative reasoning. Finally, integration of AWE and AI feedback was conducive to enriching the views or structural patterns of

students at different levels, which is conducive to stimulating their original arguments or creations. For example, a high-level student said, "After ChatGPT's suggestions, I had my own ideas. I connected different views to form an innovative new view." This confirmed that integration of AWE and AI feedback was conducive to improving students' creativity. Therefore, AWE feedback clears the language barrier for students to engage in high-order thinking, while AI feedback focuses more on the development of high-order thinking. The integration of AWE and AI feedback has a positive effect on students' development of high-order cognition.

Theme 3: Changes in Revision Strategies and Thinking Processes. The interview questions are highly consistent with students' strategic revision awareness, logical reconstruction, and metacognitive reflection. Integration of AWE and AI feedback can help students improve their revision awareness and methods. In the past, revisions only focused on surface errors, but now not only surface errors need to be revised, but also higher-order thinking (HOT) needs to be emphasized. For example, a high-level student said, "After Pigai's feedback, I will first revise the language. After ChatGPT's feedback, I will think about the argument, the validity of the argument, and the rationality of the structure." This proved that students had strategic thinking in the revision process, which is a sign of higher-order thinking (HOT). In addition, integration of AWE and AI feedback was conducive to the logical reorganization and thinking reorganization of students at different levels. For example, a middle-level student said, "ChatGPT will raise issues such as clarity, coherence, and logic in my writing. I am aware of these problems and make revisions, and finally present a good writing." Finally, the integration of AWE and AI feedback made students at different levels understand that argumentative writing is a continuous thinking process. For example, a low-level student said, "I have never thought critically about the problems in ChatGPT feedback before, but now I am used to the happiness that this higher-order thinking brings me in writing." Therefore, integration of AWE and AI feedback has realized the transformation of students from surface revision to deep revision, which is the core sign of high-order thinking.

Theme 4: Challenges and Future Preferences. The interview questions were mainly conducted from the perspectives of challenges, dependence, and future suggestions. A small number of low-level students found it difficult to understand some of the AI feedback, while the AWE feedback focused on the surface of the language. For example, a low-level student said, "I sometimes don't know which suggestion about argument structure to

choose from ChatGPT, but AWE can't provide me with clear instructions." This confirmed that low-level students face difficulties in the feedback process and may need the intervention of a teacher. In addition, some students are worried that AI feedback may be too mechanical and repetitive, and over-reliance on it may lead to a decline in their independent thinking ability. For example, a middle-level student said, "ChatGPT feedback tends to increase my laziness, which may reduce my thinking." This confirmed that over-reliance on ChatGPT may weaken students' higher-order thinking (HOT). Finally, students had a positive attitude toward the application of the integration of AWE and AI feedback, but it would be better to have teacher feedback. For example, a low-level student said, "I hope the teacher can explain some of the feedback from AWE and AI, which will help me think more clearly." This further confirmed that teachers cannot be completely replaced by technology, and the combination of technology and teachers in the future will be more conducive to students' development.

Discussion

This study employed a mixed-methods approach to explore the impact of combining AWE with AI (Artificial Intelligence) feedback on higher-order thinking (HOT) essay writing among Chinese EFL undergraduates. First, quantitative results showed that the combination of AWE and AI feedback significantly improved HOT (Analysis, Evaluation, and Creativity) in the experimental group (EG), with significantly better results than AWE feedback alone in the control group (CG). This indicates that the combination of AWE and AI feedback is more effective in enhancing students' deeper cognitive abilities in writing. The experimental group (EG) had a larger effect size (Cohen's $d > 0.80$) and showed the greatest improvement in Analysis, Evaluation, and Creativity, significantly outperforming the control group (CG). This further confirms that AWE tools can help students reduce surface language errors (Fan, 2023), while AI can provide learners with feedback on argumentation, evidence, and structure, thereby enhancing students' higher-order thinking (HOT; Mahapatra, 2024; Suh et al., 2025). Escalante et al. (2023) found that AI feedback can improve students' writing quality, and this study further extends this to the positive effect of combining AWE and AI on higher-order thinking (HOT). The integration of AWE and AI feedback provides students with dual support from both surface-level and higher-order feedback. This finding validates a revised version of Bloom's Taxonomy (Anderson & Krathwohl, 2001), which outlines the transition of students from lower-order cognitive processes

(memory and comprehension) to higher-order cognitive processes (analysis, evaluation, and creation). This research also supports cognitive feedback theory (Shute, 2008), suggesting that the integration of AWE and AI feedback encourages students to internalize feedback and strategically modify and adjust it. In conclusion, the integration of AWE and AI feedback validates cognitive feedback theory and Bloom's hierarchy of cognitive framework, extending its application to the field of human-computer collaborative learning.

Qualitative research results indicate that students perceive AWE feedback as helpful in correcting surface errors, while AI feedback helps improve their higher-order thinking (HOT; analysis, evaluation, and creation). This view aligns with the revised Bloom's Taxonomy of Cognitive Objectives (Anderson & Krathwohl, 2001), which posits that learners progress from lower-order cognition (language level) to higher-order processes (analysis, evaluation, and creation). The combination of these two feedback plays a crucial role in recursive reflection and transitions between different levels, suggesting that it can serve as a cognitive scaffold to facilitate deeper learning. Shafiee Rad (2025) found that AI enhances learners' self-regulated learning by prompting them to monitor, evaluate, and adjust their cognitive processes. This study confirms this, showing that the combination of AWE and AI encourages students to actively analyze, evaluate, and create argumentative essays. Students' progression from surface correction to higher-order thinking (HOT) modification demonstrates that cognitive feedback mechanisms support the internalization of feedback and the use of autonomous strategies. This confirms cognitive feedback theory (Shute, 2008), which states that effective feedback helps guide students' self-regulated learning. Therefore, the interaction between AWE and AI feedback promotes students' self-regulated learning behavior and enhances their feedback literacy. However, students also face difficulties in understanding feedback content and over-reliance on feedback, indicating an imbalance in the development of students' self-regulated learning, which may lead to a decline in higher-order thinking (HOT). Students hope that future teacher feedback can be integrated into the entire process of technology-enhanced feedback, which will help them understand different feedback methods. This finding confirms the positive impact of AI on students' higher-order reasoning abilities (Kim et al., 2024). This suggests that students want scaffolding to improve their self-regulated learning. In summary, these findings extend cognitive feedback theory and Bloom's taxonomy, clarifying how different feedback methods can facilitate students' transition from linguistic accuracy to higher-order thinking (HOT). In future teaching, teachers can focus on cultivating students' feedback literacy

and self-regulated learning to maximize the cognitive advantages of technology-enhanced feedback.

Conclusion, Implication, and Recommendation

This study has two main findings: Quantitative research results showed that the integration of AWE and AI feedback can improve students' higher-order thinking (HOT; analysis, evaluation, creation) in argumentative writing more than AWE alone. Qualitative data further confirmed this result, showing that students believed that integration of AWE and AI feedback helped improve their higher-order thinking (HOT), but students also faced challenges in the process.

This study has significant theoretical, pedagogical, research, and practical implications for EFL/ESL writing instruction. Firstly, based on a revised Bloom's Taxonomy and cognitive feedback theory, it fills a gap in research on human-computer collaboration in students' higher-order thinking (HOT) in argumentative writing. In addition, it provides teachers with new ideas for integrating technology into their argumentative writing instruction to address students' challenges in both surface language and deep thinking. This research is crucial for teachers' instructional design and curriculum development; courses should be iterative and feedback differentiated to improve learning outcomes for students at different levels. Teachers also need to actively participate in various training programs to enhance their teaching skills. Moreover, this study provides a reference for students to utilize technology to improve their higher-order thinking (HOT; analysis, evaluation, and creation) in argumentative writing. It guides students to enhance their self-regulated learning, critically evaluate feedback content, and cultivate their awareness and ethical consciousness regarding AI tools. Finally, this study provides a new research direction for future research—higher-order thinking (HOT)—which needs to further adopt diverse feedback methods within an ethical framework. This research also inspires future technological feedback to provide different types and depths of feedback based on learner characteristics or task complexity. In conclusion, this research has significant implications in terms of theory, teaching, research, and practice.

The study is not without limitations. For one, the sample size of this study is too small and only focuses on 64 English majors in a public university in China. Future research can be conducted on students from different regions, different colleges, different majors, and different disciplines. This study only conducted a 16-week experiment, but high-order thinking (HOT) is a continuous process, and future research can focus on its long-term impact. Moreover, this study only used AWE (Pigai) and AI (ChatGPT) tools, and future research can use more

diverse platforms. Future research can use diverse research methods such as eye tracking or thinking aloud. Future research can also add teacher feedback on the basis of technical feedback to achieve the maximum effect of human-computer collaboration. Furthermore, over-reliance on AI may limit students' independent thinking and weaken their higher-order thinking (HOT; analysis, evaluation, and creation). The application of AI in writing instruction may raise issues of academic integrity, data privacy, and educational equity. Future researchers need to carefully consider these factors to ensure the effective integration of AI into writing instruction.

Acknowledgments

The authors would like to thank all the participants who participated in this study.

ORCID iDs

Hongxia Hao  <https://orcid.org/0009-0000-2832-3272>

Abu Bakar Razali  <https://orcid.org/0000-0002-3181-1004>

Ruijia Zuo  <https://orcid.org/0000-0001-9365-5249>

Ethical Considerations

This study has received ethical approval from the target university's ethics committee. All procedures involving human subjects were conducted in accordance with established procedures and their subsequent revisions. Informed consent was provided by all participants before participation, and all data has been anonymized to protect privacy.

Consent to Participate

All participants were fully informed of the study's purpose, procedures, potential risks, and benefits. All participants signed written informed consent before participation and may withdraw from the study at any time without penalty.

Author Contributions

Hao Hongxia: Conceptualization, Investigation, Methodology, Resources, Formal analysis, Methodology, Project administration, Supervision, Writing—original draft, Writing—review & editing. Abu Bakar Razali: Review & editing, Supervision. Zuo Ruijia: Review & editing, Supervision.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

Data sets generated and analyzed in this study can be obtained from the corresponding author if there is a reasonable need.

References

- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: Complete edition*. Addison Wesley Longman, Inc.
- Borge, M., Smith, B. K., & Aldemir, T. (2024). Using generative AI as a simulation to support higher-order thinking. *International Journal of Computer-Supported Collaborative Learning*, 19(4), 479–532. <https://doi.org/10.1007/s11412-024-09437-0>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20(1), Article 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education*, 20(1), Article 57. <https://doi.org/10.1186/s41239-023-00425-2>
- Fan, N. (2023). Exploring the effects of automated written corrective feedback on EFL students' writing quality: A mixed-methods study. *SAGE Open*, 13(2), 21582440231181296. <https://doi.org/10.1177/21582440231181296>
- Fu, H., & Liu, X. (2022). EFL learner engagement in automatic written evaluation. *Frontiers in Psychology*, 13, Article 871707. <https://doi.org/10.3389/fpsyg.2022.871707>
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language teaching*, 39(2), 83–101. <https://doi.org/10.1017/S0261444806003399>
- Janse van Rensburg, J. (2024). Artificial human thinking: ChatGPT's capacity to be a model for critical thinking when prompted with problem-based writing activities. *Discover Education*, 3(1), Article 42. <https://doi.org/10.1007/s44217-024-00113-x>
- Kim, D., Majdara, A., & Olson, W. (2024). A pilot study inquiring into the impact of ChatGPT on lab report writing in introductory engineering labs. *International Journal of Technology in Education*, 7(2), 259–289. <https://doi.org/10.46328/ijte.691>
- Lee, J., & Lee, J. (2024). Development of argumentative writing ability in EFL middle school students. *Reading & Writing Quarterly*, 40(1), 36–53. <https://doi.org/10.1080/10573569.2022.2161438>
- Li, M., Li, C., Chen, Y., & Hwang, G. J. (2025). Effects of an automated evaluation mechanism on students' writing performance and higher-order thinking in an AR-based formative-peer-assessment learning mode. *Education and Information Technologies*, 30, 21889–21928. <https://doi.org/10.1007/s10639-025-13611-8>
- Li, Y., Nikitina, L., & Riget, P. N. (2022). Development of syntactic complexity in Chinese university students' L2 argumentative writing. *Journal of English for Academic Purposes*, 56, Article 101099. <https://doi.org/10.1016/j.jeap.2022.101099>
- Liu, Y., Storch, N., & Morton, J. (2022). It takes two to tango: Investigating teacher-student interactions related to written corrective feedback with Activity Theory. *Assessing Writing*, 53, Article 100647. <https://doi.org/10.1016/j.asw.2022.100647>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11(1), Article 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Mokhtar, M. M., Jamil, M., Yaakub, R., & Amzah, F. (2020). Debate as a tool for learning and facilitating based on higher order thinking skills in the process of argumentative essay writing. *International Journal of Learning, Teaching and Educational Research*, 19(6), 62–75. <https://doi.org/10.26803/ijlter.19.6.4>
- Naznin, K., Al Mahmud, A., Nguyen, M. T., & Chua, C. (2025). ChatGPT Integration in higher education for personalized learning, academic writing, and coding tasks: A systematic review. *Computers*, 14(2), Article 53. <https://doi.org/10.3390/computers14020053>
- Nunes, A., Cordeiro, C., Limpo, T., & Castro, S. L. (2022). Effectiveness of automated writing evaluation systems in school settings: A systematic review of studies from 2000 to 2020. *Journal of Computer Assisted Learning*, 38(2), 599–620. <https://doi.org/10.1111/jcal.12635>
- Pei, Z., Zheng, C., Zhang, M., & Liu, F. (2017). Critical thinking and argumentative writing: Inspecting the association among EFL learners in China. *English Language Teaching*, 10(10), 31–42. <http://doi.org/10.5539/elt.v10n10p31>
- Peng, C., & Bao, Z. (2023). Effects of reasoning demands triggered by genre on Chinese EFL learners' writing performance. *Frontiers in Psychology*, 14, Article 1164262. <https://doi.org/10.3389/fpsyg.2023.1164262>
- 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>
- Shafiee Rad, H. (2025). Reinforcing L2 reading comprehension through artificial intelligence intervention: Refining engagement to foster self-regulated learning. *Smart Learning Environments*, 12(1), Article 23. <https://doi.org/10.1186/s40561-025-00377-2>
- Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*, 36(2), 187–209. <https://doi.org/10.1017/S0958344023000265>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Solikhah, N. A. (2023). The impact of technology in teaching and learning English as foreign language: TESOL context. *Journal Corner of Education, Linguistics, and Literature*, 3(1), 83–91. <https://doi.org/10.54012/jcell.v3i1.194>

- Suh, S., Bang, J., & Han, J. W. (2025). Developing critical thinking in second language learners: Exploring generative AI like ChatGPT as a tool for argumentative essay writing. *arXiv preprint arXiv:2503.17013*. <https://doi.org/10.48550/arXiv.2503.17013>
- Wambsganss, T., Janson, A., & Leimeister, J. M. (2022). Enhancing argumentative writing with automated feedback and social comparison nudging. *Computers & Education*, 191, Article 104644. <https://doi.org/10.1016/j.compedu.2022.104644>
- Wu, M., Duan, F., & Ma, Z. (2023). Macro-Structure of argumentative text as manifestation of Chinese EFL students' critical thinking skills. *Adult and Higher Education*, 5(1), 48–59. <https://doi.org/10.23977/aduhe.2023.050107>
- Yang, H., Gao, C., & Shen, H. Z. (2024). Learner interaction with, and response to, AI-programmed automated writing evaluation feedback in EFL writing: An exploratory study. *Education and Information Technologies*, 29(4), 3837–3858. <https://doi.org/10.1007/s10639-023-11991-3>