

The Effects of ChatGPT Feedback on College Students' Argumentative Writing Self-Efficacy and Writing Performance

Chenlong Zhang

Faculty of Arts and Social Sciences, The University of Sydney, Sydney 2006, Australia

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: Generative AI tools such as ChatGPT are increasingly integrated into the teaching of English writing, and previous research confirms their ability to provide personalized feedback to correct grammatical and lexical errors and improve writing quality. However, few empirical studies have examined how ChatGPT shapes students' argumentative writing self-efficacy. To fill this research gap, a circular framework of writing-feedback-revision was used for three weeks in this study. Forty-five Chinese English majors were divided into a high-level group ($n = 22$) and a low-level group ($n = 23$) according to their vocabulary test scores, and then randomly assigned to the ChatGPT feedback group and the teacher feedback group. The research instruments included an argumentative writing assignment, a writing self-efficacy questionnaire, and a vocabulary assessment. This paper examines the effects of feedback at different levels of writing on writing competence and self-efficacy. The results reveal three key findings. First, learners who received teacher feedback (mean rank = 27.39) made greater writing progress than those who used ChatGPT (mean rank = 19.19), with low-proficiency students showing the weakest improvement with AI feedback (mean rank = 13.63). Second, ChatGPT significantly improved the writing self-efficacy of low-level learners, but there was no corresponding gain in their writing performance. Third, teacher feedback had a consistent positive effect on both writing outcomes and self-efficacy. The evidence suggests that ChatGPT cannot completely replace writing instructors, especially for low-proficiency learners, as over-reliance on AI may inhibit writing growth. Nevertheless, ChatGPT can still be used as a viable aid to correct errors and refine expressions to develop writing confidence, and advanced learners need teacher-led training to develop critical and creative thinking. This study provides empirical evidence for AI-enhanced writing instruction and provides differentiated instructional designs for learners with different proficiency levels.

Keywords: ChatGPT; Argumentative writing; Writing self-efficacy; Teacher feedback

Online publication: August 6, 2026

1. Introduction

ChatGPT is a generative artificial intelligence developed by OpenAI, which can realize human-computer interactive question answering, text translation, and output personalized modification feedback. Applying

such artificial intelligence techniques to English learning scenarios has become a new approach to language acquisition ^[1]. AI-assisted writing tools can improve learners' writing quality by correcting grammatical and lexical errors in their texts ^[2]. In addition, such intelligent tools can also provide suggestions for word substitution and sentence optimization ^[2,3]. Mahapatra ^[4] proposed that artificial intelligence tools represented by ChatGPT have a positive effect on the improvement of students' writing skills and writing ability, and their influence on English writing teaching may be more prominent in the future.

When learners use ChatGPT for writing assistance, they can set different instructions for the model to evaluate their own writing, including generating new writing ideas, providing idiom expressions, and adapting sentence pattern references ^[5]. This kind of automated and targeted feedback can guide learners to independently explore and solve writing problems, expand creative ideas, and optimize learning outcomes ^[6]. In addition, ChatGPT can also be used as writing materials to help learners supplement their background knowledge and inspire their writing ^[7].

Argumentative writing is one of the most difficult tasks for students in higher education ^[8]. It has high requirements for students' argumentation ability, and also needs to have enough knowledge and confidence, so as to make the argument more convincing. Therefore, it is both appropriate and feasible to test the effect of ChatGPT on students' writing with the task of argumentative writing ^[9,10]. Rahman and Watanobe ^[11] pointed out that if learners rely too much on ChatGPT when creating texts, they may lose the opportunity to exercise critical thinking, and it is difficult to judge whether the information given by ChatGPT is reliable. In addition, Farhi *et al.* ^[12] also proposed that students may use ChatGPT to cheat in online examinations. Therefore, the introduction of ChatGPT into the field of education may pose a threat to the traditional education system.

In order to find out whether students can deal with argumentative writing more efficiently and confidently after using ChatGPT, this study also measured learners' writing self-efficacy. The so-called writing self-efficacy refers to whether students have confidence in their writing ability, and whether they have a clear understanding of what is strong and what is weak in their writing, which is precisely a key factor in the development of writing ability ^[13,14]. Students with strong self-efficacy will have greater confidence and firmer belief in English writing. In other words, this kind of students can take the initiative to face and solve writing problems, continue to hone their writing skills, and eventually produce higher-quality exercises. Under the traditional teaching mode, the feedback given by teachers for students' writing can effectively cultivate learners' writing self-efficacy. Relying on teachers' revision suggestions and positive incentives, students can express their own views and emotions more actively ^[15]. However, as mentioned earlier, artificial intelligence tools like ChatGPT also have the ability to provide personalized feedback to the learner. ChatGPT can output immediate writing assistance feedback, which not only improves students' writing level and language ability, but also is expected to enhance their writing self-efficacy ^[16].

At present, research related to ChatGPT's writing application mostly focuses on the influence of artificial intelligence tools on writing output, such as the effect of artificial intelligence on students' writing performance ^[17] and ChatGPT in practice when applied to English essay writing ^[18]. Existing research has explored the change effect of ChatGPT's assisted writing mode on students' learning outcomes, covering the double improvement of learners' writing ability and writing skills. However, there are very few studies on whether the use of ChatGPT in the writing process can effectively enhance students' writing self-efficacy. Only a few scholars have examined the effect of using ChatGPT on writing motivation and autonomy ^[16,19]. Moreover, these studies do not distinguish writing styles from learners' English proficiency. For example,

Kızıldaş^[6] studied the writing self-efficacy of primary school students and pointed out that the use of ChatGPT can promote their creative writing skills and self-efficacy. However, limited by the small sample size and the immature thinking of primary school students, it is difficult for that study to show that ChatGPT can have an effect on the self-efficacy of all learners. Therefore, this study intends to examine the impact of ChatGPT on learners' writing self-efficacy and writing performance in argumentative writing through controlled experiments, and then provide some empirical evidence for writing teaching and practice.

2. Literature review

2.1. Writing self-efficacy

Writing self-efficacy refers to an individual's judgment of whether he or she is capable of completing a writing task^[14]. In other words, writing self-efficacy is related to learners' subjective evaluation of their own ability to complete writing tasks and their individual writing confidence level^[20]. This efficacy perception can affect learners' writing goal planning, the time and energy invested in writing, and the emotional state generated in the writing process^[20]. It can be seen that learners' writing self-efficacy can play a role in predicting their subsequent learning performance and writing achievement^[21,22].

Many empirical studies have shown that learners with a high level of writing self-efficacy tend to show more positive learning behaviors and achieve better writing scores^[21,23,24]. Relevant studies have confirmed that learners with high writing self-efficacy are more willing to take the initiative to challenge difficult writing tasks, have stronger persistence in the face of language expression and content conception problems, and use more self-regulated learning strategies^[21,23,24]. Similarly, other scholars have proposed that learners with high writing self-efficacy have a lower probability of negative emotions such as anxiety^[25]; they will show stronger learning motivation and adopt more effective writing strategies^[23,26]. At the same time, current studies also state that there is a positive correlation between writing self-efficacy and writing motivation^[27–29], and it is negatively correlated with writing anxiety and negative emotions^[30,31]. For example, Li^[30] conducted a study among 595 English major undergraduates in a Chinese university and found that high writing self-efficacy can enhance writing motivation, thereby reducing learners' academic pressure and negative emotions.

In the context of the step-by-step integration of artificial intelligence into language learning and writing, some scholars have begun to explore the impact of AI on writing self-efficacy. Many studies have found that AI-assisted writing can effectively improve writing self-efficacy^[32–34]. For example, Pellas^[33] conducted a study on 64 Greek college students and found that with the help of generative artificial intelligence, students' writing self-efficacy has been significantly improved. Similarly, Liang *et al.*^[35] pointed out in their study that learners' writing self-efficacy can be effectively enhanced through AI functions such as grammar checking, writing style modification, and providing personalized feedback. However, Zhang *et al.*^[36] mentioned in the study that not all learners can improve their writing self-efficacy with the help of artificial intelligence. For learners with high language proficiency, the use of AI was found not to be related to the improvement of writing self-efficacy. As revealed by Lin and Wang^[37], high writing self-efficacy inhibits learners' dependence on generative artificial intelligence, which indicates that generative artificial intelligence does not necessarily improve the writing self-efficacy of learners with strong language and logical structure ability. Similarly, Zhang and Xu^[38], in a study of 348 Chinese college students, found that although AI improved students' writing self-efficacy, it also deepened their dependence on AI. Once they lost AI support, it could

lead to a significant decline in learning efficiency and confidence in completing tasks. Therefore, it is not clear whether AI can actually promote learners' learning efficiency and confidence in writing activities.

2.2. Argumentative writing

Argumentative writing is a complex task for second language learners, as learners rely on a coherent logical context and convincing evidence to provide a solid argument^[39,40]. They need to articulate their claims, arguments, rebuttals, counterarguments, and conclusions in turn, as well as to use rhetorical strategies and specific expressive routines to construct claims and arguments^[41,42]. It is precisely because of the complexity of argumentative writing that learners often encounter many challenges, such as the lack of background knowledge to support the argument, and the inability to collect enough information and conduct in-depth argumentation in a limited time^[43,44].

As generative AI, such as ChatGPT, is increasingly used as a language learning tool in argumentative writing, a growing number of scholars have investigated its impact on argumentative writing. On the one hand, some researchers have observed that the use of generative AI can help learners develop their logic. For instance, Guo and Lee^[45] found in their empirical study that using generative AI can enhance learners' logical learning. Similarly, Zhang *et al.*^[46] pointed out, in a study of 15 Chinese EFL learners, that generative AI has the potential to correct learners' logical fallacies and help them organize their writing logic. On the other hand, existing research has shown that the use of generative AI in argumentative learning and writing also affects learners' critical thinking. Minh^[47] and Hikmah & Walida^[48] both noted that generative AI plays a significant role in improving learners' critical thinking skills. Emran *et al.*^[49] also found in their study that learners achieved higher scores in critical thinking after using generative AI to assist their writing. Furthermore, an experiment by Guo *et al.*^[50] supported the finding that using generative AI can help learners improve their argumentation skills during classroom debates and argumentative writing tasks.

In addition, research work has shown that the use of generative artificial intelligence in argumentative learning and writing can also have an effect on learners' critical thinking. Minh^[47] and Hikmah and Walida^[48] have pointed out that generative artificial intelligence plays an important role in improving learners' critical thinking ability. Emran *et al.*^[49] also found in the study that learners obtained higher scores in critical thinking after using generative artificial intelligence to assist writing. In addition, an experiment conducted by Guo *et al.*^[50] also supported the finding that the use of generative artificial intelligence can help learners improve their argumentation ability in classroom debate and argumentative writing tasks.

Although these studies generally tend to support that generative artificial intelligence can help learners and show better critical thinking and logic in argumentative writing, there are still some studies that believe that the impact of generative artificial intelligence on argumentative writing is not so clear. For example, Liang and Wu^[51] found in a study of 58 students at a university in Macao that the use of AI did not bring about a particularly significant change in learners' critical thinking. Similarly, Zhang *et al.*^[52] also mentioned that most of the previous studies in this field mainly relied on semi-structured interviews and questionnaires to collect students' and teachers' opinions, which made the research results more subjective and uncertain, and lacked sufficient evidence to prove that AI really improved learners' writing skills and critical thinking ability.

3. Research question

Based on the literature review above, current research on ChatGPT-assisted writing has primarily focused on its effects on learners' writing performance, with relatively little exploration of its role in writing self-efficacy.

In addition, the existing research has not been able to distinguish the learners' writing level or writing genre well. Therefore, this study intends to use a three-week controlled experiment to specifically look at the impact of ChatGPT feedback on the performance of argumentative writing and writing self-efficacy of college students with different English levels. There are several questions to be explored:

- (1) Compared with teacher feedback, will ChatGPT feedback have a more significant effect on the performance of argumentative writing of college students at different English levels (including high and low levels)?
- (2) Compared with teacher feedback, can ChatGPT feedback produce a more significant promotion effect on the writing self-efficacy of college students with different English levels (high and low)?
- (3) Which is more effective to improve learners' writing self-efficacy and writing performance, teacher feedback or ChatGPT feedback?

4. Methods

4.1. Participants

This study recruited 45 Chinese college students to participate in the experiment through online registration. Throughout the process, the study fully respected the rights of each participant. All participants had a clear understanding of the purpose and specific process of this study before the start of the experiment, and also clearly knew that they could choose to withdraw from the study at any time without any impact.

All 45 students who participated in the experiment were from English-related majors, including 37 females and 8 males. Before the experiment began, each participant signed an informed consent form and was told that all the tests needed to be completed, and that no search tools were allowed and no cheating was allowed during the test. Moreover, after confirming their participation in the experiment, participants were also asked to fill out a questionnaire to collect basic information, including gender, the frequency of ChatGPT use (such as several times a week or every day), and the frequency of English writing (once a week, twice a week, or more).

4.2. Instruments

4.2.1. Vocabulary test

Learners' vocabulary level is widely regarded as the foundation of English writing and reflects learners' writing development^[53]. Nation^[54] has pointed out that learners' writing proficiency can be assessed by measuring their vocabulary level. In addition, a large number of studies have also shown that vocabulary proficiency tests can measure and judge learners' writing performance to a certain extent^[55-58]. Based on this, this study will use vocabulary tests to evaluate the writing proficiency of all participants, and then divide them into a high-level group and a low-level group according to the results of vocabulary tests.

Before the experiment, all participants completed an online vocabulary test which took about 20 to 30 minutes (<https://www.lexutor.ca/tests/levels/productive/>). This test is a very reliable and effective tool to assess learners' vocabulary growth and English proficiency, and it helps to divide learners into different

levels of 2k, 3k, and 5k ^[59]. In addition, the test is particularly suitable for measuring the English proficiency of foreign language learners ^[60], so it is very suitable for the participants of this study, that is, Chinese foreign language learners.

The test content is the 2k-word and 3k-word level selected from the Vocabulary Test Version A. After completing the vocabulary test, the students were assigned to the corresponding groups according to their scores: 22 in the high-level group and 23 in the low-level group. After data screening, it was found that all participants' scores actually fell within the range of 2k to 3k levels, so no participants needed to enter the 5k vocabulary level test.

4.2.2. Self-efficacy scale

This study used a writing self-efficacy questionnaire to measure participants' writing self-efficacy. This questionnaire has a total of 20 items, which is adapted from the Questionnaire of English Writing Self-Efficacy by Sun and Wang ^[61]. The questionnaire divides writing self-efficacy into five core dimensions. They are: idea generation (e.g., "I can think of many ideas for writing"), organization (e.g., "I can organize sentences into paragraphs to express opinions"), language mechanics (e.g., "I can spell all words correctly in writing"), grammar and syntax (e.g., "I can use verb tenses correctly in English writing"). There is also self-regulation (e.g., "I can concentrate on writing for at least an hour"). The scoring method uses a 7-point Likert scale, and the score ranges from 1 point (representing "I can't do it at all") to 7 points (representing "I can do it at all"). In this study, the Cronbach's α coefficient of the scale reached 0.916, which indicates that it has a fairly good internal consistency reliability.

4.3. Procedure

The whole experiment lasted for three weeks, including three writing tasks, using a cyclic design of "writing-feedback-revision." The specific process is shown in **Figure 1** below:

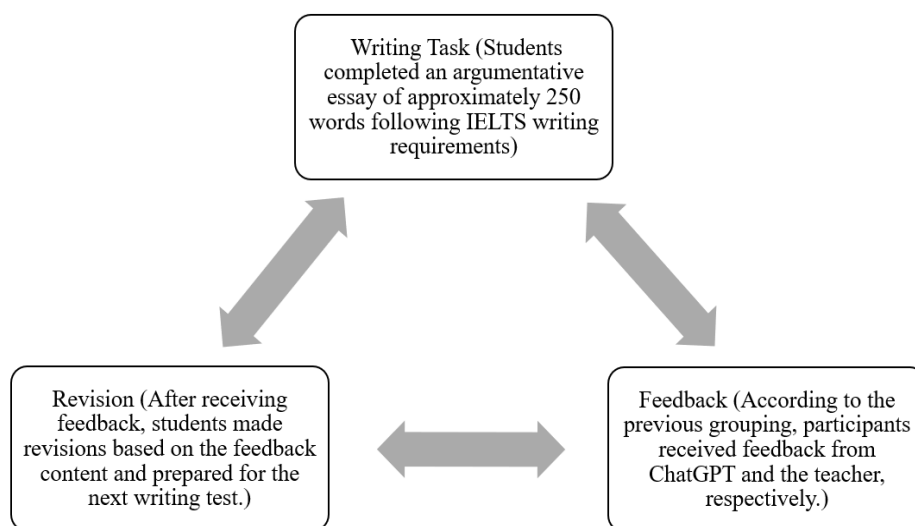


Figure 1. Writing-feedback-revision process

4.3.1. Grouping

Before the experiment, 45 participants were divided into a high-level group ($n = 22$) and a low-level group ($n = 23$) according to their vocabulary test scores. Then, within each group, 50% of the participants were randomly assigned to the teacher feedback group, and the remaining 50 % were assigned to the ChatGPT feedback group. In this way, the study formed four groups: ChatGPT feedback + high-level group ($n = 12$), ChatGPT feedback + low-level group ($n = 12$), teacher feedback + high-level group ($n = 10$), and teacher feedback + low-level group ($n = 11$).

4.3.2. Three-round writing task

All participants took a 45-minute online writing test at the same time every week. The writing tasks and requirements were selected from Task 2 (argumentative writing) in the IELTS writing question bank. In order to ensure fairness and the validity of the experimental data, all participants were given the same writing task. In the process of answering questions, they were asked to record the whole computer screen and their hand position with additional equipment (mobile phone or computer), which was convenient for researchers to monitor the test process.

As soon as the writing task is completed, the compositions of the participants in the teacher feedback group will be randomly sent to two teachers for independent scoring. The scoring criteria are based on the framework of Jacobs^[62]. The total score is set at 100 points, which is divided into five dimensions: Content 30 points, Organization 20 points, Vocabulary 20 points, Language Use 25 points, and Mechanics 5 points. If the scores of the two teachers are too different, then ask a third teacher to evaluate the composition, and finally determine a final score after discussion. In addition to scoring, teachers will also give suggestions on the five aspects of content, organization, vocabulary, language use, and mechanics. In order to ensure that the score of each composition is consistent, the intra-class correlation coefficient (ICC) is calculated after every 10 compositions are evaluated. Once the ICC value is found to fall below 0.8, the scoring team has to recalibrate the scoring criteria before scoring.

While the teacher writes feedback to the participants, the researcher uses the same ChatGPT account (using the GPT-4 model) to provide revision suggestions for the participants' compositions in the ChatGPT feedback group. The prompts input into ChatGPT are standardized as follows: "As a writing teacher, please provide detailed feedback on this thesis. Focus on Content, Organization, Vocabulary, Language Use, and Mechanics. Please provide detailed suggestions on each aspect in a friendly tone." In other words, ChatGPT and teachers actually provide students with revision suggestions and feedback in the same dimension. After receiving feedback from ChatGPT or teachers, participants will revise their compositions according to the feedback, and consciously apply the points learned in the previous round of feedback when completing the follow-up writing tasks.

5. Data analysis

All the quantitative data were analyzed using SPSS 27.0 software. The Cronbach's α coefficient was calculated to evaluate the internal consistency reliability of the writing self-efficacy scale. In addition, descriptive statistics, including mean (M) and standard deviation (SD), were also calculated to describe the characteristics of the distribution of the four groups of participants on each variable.

As the Kolmogorov-Smirnov normality test indicated that writing improvement scores, pre-test self-

efficacy scores, post-test self-efficacy scores, and self-efficacy improvement scores did not follow a normal distribution ($P < 0.05$), non-parametric tests were primarily employed for subsequent analyses. Specifically, the Kruskal-Wallis H test was used to compare the differences among the four groups in writing improvement scores and self-efficacy improvement scores. To further examine within-group effects, the Wilcoxon signed-rank test was used to assess changes in each group before and after the experiment. For the writing pre-test and post-test data, which followed a normal distribution, independent samples t -tests and paired samples t -tests were conducted as supplementary analyses.

In the analysis, writing proficiency (high-level/low-level) and feedback type (ChatGPT feedback/teacher feedback) were included as two independent variables to examine their main effects and interaction effects on the dependent variables. Additionally, the relative performance of the four groups in writing improvement and self-efficacy enhancement was visually presented by comparing the mean ranks of the improvement scores. The significance level α for all statistical tests was set at 0.05.

5.1. Normality test

Before conducting the formal analyses, this study first performed normality tests on each dependent variable. The Kolmogorov-Smirnov (K-S) test was used to examine whether writing pre-test (writing1), writing post-test (writing3), writing improvement score (write_gain), self-efficacy pre-test (pre_mean), self-efficacy post-test (post_mean), and self-efficacy improvement score (gain_mean) followed a normal distribution. The test results are presented in **Figure 2**.

Variable	Sample Size (N)	Mean (M)	Standard Deviation (SD)	K-S Statistic	P	Normality (Yes/No)
writing1	45	74.49	8.193	0.130	0.053	Yes
writing3	45	77.73	8.389	0.090	0.200	Yes
write_gain	45	3.24	7.598	0.149	0.013	No
pre_mean	45	5.22	0.791	0.138	0.032	No
post_mean	45	5.54	0.769	0.133	0.045	No
gain_mean	45	0.32	0.835	0.132	0.047	No

Figure 2. Kolmogorov-Smirnov (K-S) test results

As shown in **Figure 2**, the K-S test significance values for the writing pre-test ($P = 0.053$) and writing post-test ($P = 0.200$) were both greater than 0.05, indicating that they met the assumption of normal distribution. Therefore, a parametric test should be used for these two variables. In contrast, the writing progress score ($P = 0.013$), self-efficacy pre-test ($P = 0.032$), self-efficacy post-test ($P = 0.045$), and self-efficacy progress score ($P = 0.047$) all had significance values less than 0.05, which indicates that these variables do not meet the assumption of normal distribution, so they have to be tested by a non-parametric test.

5.2. Descriptive statistics

As shown in **Figure 3**, the descriptive statistics (in the form of $M \pm SD$) of the four groups of participants in writing performance and self-efficacy are clearly listed.

Group	n	Writing Pre-test	Writing Post-test	Writing Improvement	Self-efficacy Pre-test	Self-efficacy Post-test	Self-efficacy Improvement
				Score			Score
ChatGPT Feedback + High-level	12	78.50±7.14	80.92±7.25	2.42±9.74	5.56±0.70	5.64±0.93	0.08±0.59
ChatGPT Feedback + Low-level	12	68.83±10.59	72.67±14.92	3.83±9.72	5.14±0.64	5.56±0.64	0.42±0.72
Teacher Feedback + High-level	10	78.50±5.38	81.50±6.64	3.00±8.49	5.42±0.35	5.72±0.55	0.30±0.67
Teacher Feedback + Low-level	11	68.18±8.49	71.55±12.87	3.36±14.14	4.75±1.11	5.24±0.88	0.49±1.27

Figure 3. Descriptive statistics (in the form of $M \pm SD$) in writing performance and self-efficacy

According to the results of descriptive statistics, participants in the four groups have improved their writing performance and self-efficacy. Among these groups, the teacher feedback + low-level group has the greatest gain in self-efficacy ($M = 0.49$), while the ChatGPT feedback + low-level group has the greatest gain in writing performance ($M = 3.8$).

5.3. Pre-test equivalence test

The pre-test equivalence test is a key link to verify whether the random grouping is effective. This study is a pre-test equivalence test from two dimensions: one dimension is to see whether grouping according to writing level (that is, high level and low level) is effective; the other dimension is to test the randomness of grouping according to the type of feedback (that is, ChatGPT feedback and teacher feedback). For these two dimensions, the researchers tested the writing pre-test (writing1) and the self-efficacy pre-test (pre_mean). Since the writing pre-test followed a normal distribution, an independent samples *t*-test was used. As the self-efficacy pre-test did not follow a normal distribution, the Mann-Whitney U test was employed. The test results are presented in **Figure 4**.

Dimension	Variable	Test Method	Group	n	Mean (M)	Standard Deviation (SD)	Mean Rank	Test Statistic	P	Result
Grouping by Writing Proficiency	writing1	Independent Samples t-Test	High-level	22	78.5	6.29	—	t=3.88	<0.001	Significant difference (p < 0.05)
			Low-level	23	68.5	9.55	—			
		Mann-Whitney U	High-level	22	—	—	27.75	U=148.50	0.018	Significant difference (p < 0.05)
			Low-level	23	—	—	18.46			
		Independent Samples t-Test	ChatGPT feedback	24	73.67	10.25	—	t=0.33	0.74	No significant difference (p > 0.05)
			teacher feedback	21	73.33	8.95	—			
Grouping by Feedback Type	pre_mean	Mann-Whitney U	ChatGPT feedback	24	—	—	25.08	U=202.00	0.255	No significant difference (p > 0.05)
			teacher feedback	21	—	—	20.62			

Figure 4. Mann-Whitney U test results

Regarding the effectiveness of grouping by writing proficiency, the independent samples t-test results showed a significant difference in writing pre-test scores between the high-level group ($M = 78.50$, $SD = 6.29$) and the low-level group ($M = 68.50$, $SD = 9.55$), $t(43) = 3.88$, $P < 0.001$. The results of the Mann-Whitney U test showed that the scores of the high-level group in the pre-test of self-efficacy (average rank = 27.75) were significantly higher than those of the low-level group (average rank = 18.46), with the statistic $U = 148.50$, $P = 0.018$. This shows that it is effective to divide students into high-level and low-level groups according to their writing level, and the high-level group does have a stronger sense of writing self-efficacy, which is consistent with the direction expected before the study.

As for whether the grouping according to the type of feedback is completely random, the results of the independent-samples *t*-test show that there is no significant difference between the AI feedback group ($M = 73.67$, $SD = 10.25$) and the teacher feedback group ($M = 73.33$, $SD = 8.95$) in the pre-test scores of writing, $t(43) = 0.33$, $P = 0.740$. At the same time, the results of the Mann-Whitney U test also showed that there was no significant difference between the AI feedback group (mean rank = 25.08) and the teacher feedback group (mean rank = 20.62) in the pre-test scores of self-efficacy, $U = 202.00$, $P = 0.255$. This shows that the participants are indeed randomly assigned to two different feedback conditions, and the experiment has good internal validity.

5.4. Comparison of differences among the four groups

This study employed the Kruskal-Wallis H test to compare the differences among the four groups in writing improvement scores and self-efficacy improvement scores.

5.4.1. Differences in writing improvement scores

The mean ranks of writing improvement scores for the four groups are shown in **Figure 5**.

Group	n	Mean Rank
ChatGPT Feedback + High-level	12	24.75
ChatGPT Feedback + Low-level	12	13.63
Teacher Feedback + High-level	10	27.95
Teacher Feedback + Low-level	11	26.82

Figure 5. Mean ranks of writing improvement scores

The Kruskal-Wallis test results for writing improvement scores are shown in **Figure 6**.

Test Statistic	Value
Kruskal-Wallis H	8.72
Degrees of Freedom	3
Asymptotic Significanc	0.033

Figure 6. Kruskal-Wallis test results for writing improvement scores

The Kruskal-Wallis H test showed a significant difference among the four groups in writing improvement scores ($H = 8.720$, $P = 0.033$). In terms of mean ranks, the Teacher Feedback + High-level group had the highest mean rank (27.95), while the ChatGPT Feedback + Low-level group had the lowest mean rank (13.63). Overall, the teacher feedback groups (High-level: 27.95; Low-level: 26.82) outperformed the ChatGPT feedback groups (High-level: 24.75; Low-level: 13.63), and low-level students showed the smallest writing improvement under the ChatGPT feedback condition.

5.4.2. Differences in self-efficacy improvement scores

The mean ranks of self-efficacy improvement scores for the four groups are shown in **Figure 7**.

Group	n	Mean Rank
ChatGPT Feedback + High-level	12	19.13
ChatGPT Feedback + Low-level	12	25.75
Teacher Feedback + High-level	10	24.1
Teacher Feedback + Low-level	11	23.23

Figure 7. Mean ranks of self-efficacy improvement scores

The Kruskal-Wallis test results for self-efficacy improvement scores are shown in **Figure 8**.

Test Statistic	Value
Kruskal-Wallis H	1.647
Degrees of Freedom	3
Asymptotic Significance	0.649

Figure 8. Kruskal-Wallis test results for self-efficacy improvement scores

The Kruskal-Wallis H test results showed that there was no significant difference among the four groups in self-efficacy improvement scores ($H = 1.647$, $P = 0.649$). Although the AI Feedback + Low-level group had the highest mean rank (25.75), the difference did not reach statistical significance.

5.5. Pre-test and post-test comparison among the four groups

In this study, paired-sample t -tests were used for writing scores, which followed a normal distribution, while Wilcoxon signed-rank tests were used for self-efficacy scores, which did not follow a normal distribution, to examine the pre-test and post-test differences within each of the four groups.

The pre-test and post-test comparison for the ChatGPT Feedback + High-level group is shown in **Figure 9**.

Variable	Test Method	Pre-test (M±SD)	Post-test (M±SD)	Test Statistic	Significance	Conclusion
Writing Score	Paired t-test	78.50±7.14	80.92±7.25	$t=0.92$	0.375	Not significant
Self-efficacy	Wilcoxon	5.56±0.70	5.64±0.93	$Z = -0.47$	0.638	Not significant

Figure 9. Pre-test and post-test comparison for the ChatGPT Feedback + High-level group

The pre-test and post-test comparison for the ChatGPT Feedback + Low-level group is shown in **Figure 10**.

Variable	Test Method	Pre-test (M±SD)	Post-test (M±SD)	Test Statistic	Significance	Conclusion
Writing Score	Paired t-test	68.83±10.59	72.67±14.92	$t=1.14$	0.278	Not significant
Self-efficacy	Wilcoxon	5.14±0.64	5.56±0.64	$Z=-2.08$	0.037	significant

Figure 10. Pre-test and post-test comparison for the ChatGPT Feedback + Low-level group

The pre-test and post-test comparison for the Teacher Feedback + High-level group is shown in **Figure 11**.

Variable	Test Method	Pre-test (M±SD)	Post-test (M±SD)	Test Statistic	Significance	Conclusion
Writing Score	Paired t-test	78.50±5.38	81.50±6.64	$t=1.79$	0.104	Not significant
Self-efficacy	Wilcoxon	5.42±0.35	5.72±0.55	$Z=-1.36$	0.173	Not significant

Figure 11. Pre-test and post-test comparison for the Teacher Feedback + High-level group

The pre-test and post-test comparison for the Teacher Feedback + Low-level group is shown in **Figure 12**.

Variable	Test Method	Pre-test (M±SD)	Post-test (M±SD)	Test Statistic	Significance	Conclusion
Writing Score	Paired t-test	68.18±8.49	71.55±12.87	$t=1.07$	0.307	Not significant
Self-efficacy	Wilcoxon	4.75±1.11	5.24±0.88	$Z=-1.78$	0.075	Marginally significant

Figure 12. Pre-test and post-test comparison for the Teacher Feedback + Low-level group

Therefore, the ChatGPT Feedback + High-level group showed no significant improvement in either writing scores ($t = 0.92$, $P = 0.375$) or self-efficacy ($Z = -0.47$, $P = 0.638$). On the contrary, the writing self-efficacy of the low-level ChatGPT feedback group was significantly improved ($Z = -2.08$, $P = 0.037$), and the scores of 10 subjects increased, and only 2 subjects decreased. However, there was no significant change in the writing scores of this group ($t = 1.14$, $P = 0.278$). The writing scores of 9 subjects decreased, and only 3 subjects increased.

There was no significant improvement in writing scores ($t = 1.79$, $P = 0.104$) and writing self-efficacy ($Z = -1.36$, $P = 0.173$) in the high-level group, and only the improvement in writing scores was close to the marginal significant level. The edge of writing self-efficacy in the group with a low level of teacher feedback was significantly improved ($Z = -1.78$, $P = 0.075$), and the scores of 8 subjects increased, and the scores of 3 subjects decreased. However, there was no significant change in writing scores ($t = 1.07$, $P = 0.307$), 7

subjects' scores increased, and 3 subjects' scores decreased.

6. Discussion

This study aimed to investigate the effects of using ChatGPT in writing on college students' English writing self-efficacy and writing performance. Through a three-week writing test, the study examined changes in writing self-efficacy and writing performance among high- and low-writing proficiency college students under two conditions: ChatGPT feedback and teacher feedback. The following discussion will address the research questions and discuss the main findings, drawing on existing literature.

First, based on the experimental results, this study found that the use of ChatGPT provided very limited assistance to students in argumentative writing. The Kruskal-Wallis H test results showed a significant difference among the four groups in writing improvement scores ($H = 8.720$, $P = 0.033$). From the average rank, the teacher feedback + high level group is 27.95, the teacher feedback + low level group is 26.82, the ChatGPT feedback + high level group is 24.75, and the ChatGPT feedback + low level group is only 13.63. In other words, the teacher feedback group had an average rank of 27.39, while the ChatGPT feedback group had an average rank of 19.19. This result shows that the writing improvement in the teacher feedback group is significantly greater than that in the ChatGPT feedback group. This difference is particularly prominent in the group of low-level learners: the group with a low level of ChatGPT feedback has the lowest average rank of writing improvement among the four groups, while the group with a low level of teacher feedback has the second highest average rank of 26.82. The above results show that it is difficult for low-level learners to effectively improve their writing ability by only relying on ChatGPT feedback. This conclusion is consistent with the conclusion of Bašić *et al.*, who also found that using ChatGPT to assist students in writing does not improve the quality of students' texts. The potential reason for this phenomenon is that the current feedback output of ChatGPT focuses on grammar error correction and complete paragraph generation, which occupies the space for learners' independent thinking and reflection, thus hindering their acquisition of writing skills and cultivation of critical thinking^[63].

In addition, although ChatGPT can provide personalized feedback, it is difficult to provide highly targeted modification suggestions combined with learners' unique writing style, and it is easy to make learners lose their own writing characteristics and authentic expression experience^[64]. As a result, ChatGPT is often difficult for learners to understand and assimilate: tweaks that don't fit their own writing style, and reference passages that are too complete can cost them a lot of extra time. More importantly, ChatGPT's feedback rarely touches on deeper issues, such as the root causes of errors or paths to critical thinking. Only relying on grammar error correction and model essay reference cannot provide comprehensive and systematic writing guidance for low-level learners.

Secondly, the data of this study showed that the writing self-efficacy of high-level learners was not improved after using ChatGPT. The writing self-efficacy of low-level learners increased significantly ($Z = -2.08$, $P = 0.037$), but their writing scores did not change significantly ($t = 1.14$, $P = 0.278$). This indicated that the improvement of self-efficacy of low-level learners was not synchronized with the progress of writing ability. As mentioned above, it is difficult for low-level learners to digest and master the modification suggestions and model texts given by ChatGPT. However, at the same time, the grammar correction and personalized feedback provided by ChatGPT will make low-level learners develop the psychology of relying

on the tool to complete writing tasks ^[32,34,35]. Kim *et al.* ^[32] also confirmed that AI can improve learners' writing confidence. However, for high-level learners with sufficient writing experience and language reserve, they are more likely to think that the content generated by ChatGPT is insufficient in depth, so it cannot improve their writing self-efficacy by using this tool ^[36].

Finally, teacher feedback has a positive effect on the self-efficacy and writing performance of high-level and low-level learners, while ChatGPT only raises the self-efficacy of low-level learners. High-level learners have basically mastered language skills, and most of their writing skills, so their focus may be more on critical thinking and creative expression. Teachers have the ability to provide more enlightening guidance and feedback according to different learners' needs, which ChatGPT cannot do. For low-level students, although ChatGPT can give some "gentle" feedback and encouragement according to the instructions, it still cannot provide emotional support, and it is difficult to give strategies that can promote students' long-term development. Moreover, in the face of a large number of error prompts generated by ChatGPT, although low-level learners can realize that there are many errors in their compositions, they cannot figure out which problems should be corrected first.

Therefore, the results of this study have some implications for writing teaching: ChatGPT cannot fully assume the role of a teacher. Using ChatGPT to assist writing alone cannot effectively improve learners' self-efficacy or writing performance. Especially for low-level learners, relying solely on ChatGPT may not be conducive to the development of their writing ability. Based on this, in the process of writing teaching and learning, ChatGPT can be used as an auxiliary tool to help learners correct errors and improve expression, especially to help low-level learners build the confidence needed to complete writing tasks step by step. In addition, it is necessary for writing teaching to adopt different strategies according to the level of students. This study shows that high-level learners' demand for ChatGPT is actually relatively weak. On the contrary, they need more guidance from teachers to improve their critical thinking and creative writing ability. Therefore, high-level learners need more support and guidance from teachers. At the same time, low-level learners can use teacher feedback to improve their writing skills. On the other hand, they can also use artificial intelligence to complete writing tasks alone, accumulate writing experience, and gradually establish confidence.

7. Conclusion

This study adopts a three-week experiment, including three rounds of "writing-feedback-modification" cycles, to compare the effects of ChatGPT feedback and teacher feedback on argumentative writing performance and writing self-efficacy of 45 Chinese English majors (divided into high-level and low-level groups). Based on the research results, ChatGPT has a limited effect on improving argumentative writing performance, and the writing progress of the teacher feedback group is significantly greater than that of the ChatGPT feedback group. Low-level learners made the least progress in writing under ChatGPT feedback, but made significant progress under teacher feedback. This shows that for learners with low writing proficiency, relying on ChatGPT may not be conducive to the development of their writing ability. Secondly, ChatGPT improves low-level learners' writing self-efficacy, but does not improve their writing performance at the same time. This enhancement of confidence has not been translated into the actual progress of writing ability. Thirdly, teacher feedback has a sustained and stable positive effect on both writing performance

and self-efficacy. Teachers can provide more insightful feedback based on learners' individual differences, while giving emotional support and suggestions for long-term development. In contrast, although ChatGPT can generate "gentle" encouragement according to instructions, it cannot understand learners' emotions and difficulties, nor can it help low-level learners to identify which problems need to be corrected first among the many error prompts.

Although some important findings have been made in this study, the following limitations still need to be acknowledged and deserve further exploration in future studies. First, the sample size of this study is 45 people in total, which meets the basic requirements of statistical analysis, but the small sample size may reduce the accuracy of the research conclusion. Secondly, the three-week intervention period was short, and it was difficult to fully reveal the long-term effect of ChatGPT on learners' writing ability and writing self-efficacy. Follow-up studies could extend the duration of the intervention (e.g., a full semester) and conduct follow-up tests after the intervention to further explore the lasting effects of ChatGPT feedback. Third, this study only uses quantitative research methods. Although this method can output objective statistical data, it is difficult to deeply explore the subjective feelings and internal cognitive processes of learners in the face of two types of feedback. In the future, a mixed research design can be used to carry out relevant discussions by combining qualitative methods such as semi-structured interviews and open-ended questionnaires.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Roll I, Wylie R, 2016, Evolution and Revolution in Artificial Intelligence in Education. *International Journal of Artificial Intelligence in Education*, 26(2): 582–599. <https://doi.org/10.1007/s40593-016-0110-3>
- [2] Zhao X, 2023, Leveraging Artificial Intelligence (AI) Technology for English Writing: Introducing Wordtune as a Digital Writing Assistant for EFL Writers. *RELC Journal*, 54(3): 890–894. <https://doi.org/10.1177/00336882221094089>
- [3] Salvagno M, Taccone FS, Gerli AG, 2023, Can Artificial Intelligence Help for Scientific Writing? *Critical Care*, 27(1): Article 75. <https://doi.org/10.1186/s13054-023-04380-2>
- [4] 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>
- [5] Alkamel MAA, Alwagieh NAS, 2024, Utilizing an Adaptable Artificial Intelligence Writing Tool (ChatGPT) to Enhance Academic Writing Skills Among Yemeni University EFL Students. *Social Sciences Humanities Open*, 10, Article 101095. <https://doi.org/10.1016/j.ssaho.2024.101095>
- [6] Kızıldaş Y, 2025, Integration of Artificial Intelligence (AI) into Primary School Students' Writing Skills: The Impact of ChatGPT on Creative Writing and Writing Self-Efficacy. *Journal of Educational Computing Research*, 63(7–8): 1717–1747. <https://doi.org/10.1177/07356331251365187>
- [7] Levine S, Beck SW, Mah C, et al., 2025, How Do Students Use ChatGPT as a Writing Support? *Journal of Adolescent Adult Literacy*, 68(5): 445–457. <https://doi.org/10.1002/jaal.1373>
- [8] Latifi S, Noroozi O, Talaei E, 2021, Peer Feedback or Peer Feedforward? Enhancing Students' Argumentative Peer Learning Processes and Outcomes. *British Journal of Educational Technology*, 52(2): 768–784. <https://doi.org/10.1111/bjet.12511>

org/10.1111/bjet.13054

- [9] Heitmann P, Hecht M, Schwanewedel J, et al., 2014, Students' Argumentative Writing Skills in Science and First-Language Education: Commonalities and Differences. *International Journal of Science Education*, 36(18): 3148–3170. <https://doi.org/10.1080/09500693.2014.962644>
- [10] Valero Haro A, Noroozi O, Biemans H, et al., 2022, Argumentation Competence: Students' Argumentation Knowledge, Behavior and Attitude and Their Relationships with Domain-Specific Knowledge Acquisition. *Journal of Constructivist Psychology*, 35(1): 123–145. <https://doi.org/10.1080/10720537.2020.1734995>
- [11] Rahman MM, Watanobe Y, 2023, ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Applied Sciences*, 13(9): Article 5783. <https://doi.org/10.3390/app13095783>
- [12] Farhi F, Jeljeli R, Aburezeq I, et al., 2023, Analyzing the Students' Views, Concerns, and Perceived Ethics About ChatGPT Usage. *Computers and Education: Artificial Intelligence*, 5, Article 100180. <https://doi.org/10.1016/j.caeai.2023.100180>
- [13] Bandura A, 1986, *Social Foundations of Thought and Action: A Social Cognitive Theory*. Prentice-Hall.
- [14] Pajares F, Valiante G, 2001, Gender Differences in Writing Motivation and Achievement of Middle School Students: A Function of Gender Orientation? *Contemporary Educational Psychology*, 26(3): 366–381. <https://doi.org/10.1006/ceps.2000.1069>
- [15] Chung HQ, Chen V, Olson CB, 2021, The Impact of Self-Assessment, Planning and Goal Setting, and Reflection Before and After Revision on Student Self-Efficacy and Writing Performance. *Reading and Writing*, 34(7): 1885–1913. <https://doi.org/10.1007/s11145-021-10186-x>
- [16] Bouzar A, El Idrissi K, Ghourdou T, 2024, ChatGPT and Academic Writing Self-Efficacy: Unveiling Correlations and Technological Dependency Among Postgraduate Students. *Arab World English Journal*, Special Issue on ChatGPT.
- [17] Sun Z, Anbarasan M, Praveen Kumar D, 2021, Design of Online Intelligent English Teaching Platform Based on Artificial Intelligence Techniques. *Computational Intelligence*, 37(3): 1166–1180. <https://doi.org/10.1111/coin.12351>
- [18] Fitria TN, 2023, Artificial Intelligence (AI) Technology in Open AI ChatGPT Application: A Review of ChatGPT in Writing English Essays. *ELT Forum: Journal of English Language Teaching*, 12, 44–58. <https://journal.unnes.ac.id/sju/index.php/elt>
- [19] Kang JG, Pyo SY, 2024, College Students' Writing Self-Efficacy in Reflective Writing Classes Utilizing ChatGPT. *Journal of Practical Engineering Education*, 16(4): 471–479.
- [20] Pajares F, 2003, Self-Efficacy Beliefs, Motivation, and Achievement in Writing: A Review of the Literature. *Reading Writing Quarterly*, 19(2): 139–158. <https://doi.org/10.1080/10573560308222>
- [21] Bai B, Wang J, Nie Y, 2021, Self-Efficacy, Task Values and Growth Mindset: What Has the Most Predictive Power for Primary School Students' Self-Regulated Learning in English Writing and Writing Competence in an Asian Confucian Cultural Context? *Cambridge Journal of Education*, 51(1): 65–84. <https://doi.org/10.1080/0305764X.2020.1778639>
- [22] Sun T, Wang C, Lambert RG, et al., 2021, Relationship Between Second Language English Writing Self-Efficacy and Achievement: A Meta-Regression Analysis. *Journal of Second Language Writing*, 53, Article 100817. <https://doi.org/10.1016/j.jslw.2021.100817>
- [23] Golparvar SE, Khafi A, 2021, The role of L2 Writing Self-Efficacy in Integrated Writing Strategy Use and Performance. *Assessing Writing*, 47, Article 100504. <https://doi.org/10.1016/j.asw.2020.100504>

- [24] Zabihi R, 2018, The Role of Cognitive and Affective Factors in Measures of L2 Writing. *Written Communication*, 35(1): 32–57. <https://doi.org/10.1177/0741088317735836>
- [25] Redifer JL, Bae CL, Zhao Q, 2021, Self-Efficacy and Performance Feedback: Impacts on Cognitive Load During Creative Thinking. *Learning and Instruction*, 71, Article 101395. <https://doi.org/10.1016/j.learninstruc.2020.101395>
- [26] Pajares F, Valiante G, Cheong YF, 2006, Writing Self-Efficacy and Its Relation to Gender, Writing Motivation and Writing Competence: A Developmental Perspective. In S. Hidi, P. Boscolo (Eds.): *Studies in Writing*, Vol. 19, 141–159, BRILL. https://doi.org/10.1163/9781849508216_009
- [27] Mendoza L, Lehtonen T, Lindblom-Ylänne S, et al., 2022, Exploring First-Year University Students' Learning Journals: Conceptions of Second Language Self-Concept and Self-Efficacy for Academic Writing. *System*, 106, Article 102759. <https://doi.org/10.1016/j.system.2022.102759>
- [28] Shen B, Bai B, Park M, 2023, Exploring Hong Kong Primary Students' English Writing Motivation: Relationships Between Writing Self-efficacy and Task Value. *Journal of Multilingual and Multicultural Development*, 44(4): 324–338. <https://doi.org/10.1080/01434632.2020.1823397>
- [29] Teng LS, Sun PP, Xu L, 2018, Conceptualizing Writing Self-Efficacy in English as a Foreign Language Contexts: Scale Validation Through Structural Equation Modeling. *TESOL Quarterly*, 52(4): 911–942. <https://doi.org/10.1002/tesq.432>
- [30] Li B, 2022, Research on Correlation Between English Writing Self-Efficacy and Psychological Anxiety of College Students. *Frontiers in Psychology*, 13, Article 957664. <https://doi.org/10.3389/fpsyg.2022.957664>
- [31] Woodrow L, 2011, College English Writing Affect: Self-Efficacy and Anxiety. *System*, 39(4): 510–522. <https://doi.org/10.1016/j.system.2011.01.017>
- [32] Kim J, Yu S, Detrick R, et al., 2025, Exploring Students' Perspectives on Generative AI-Assisted Academic Writing. *Education and Information Technologies*, 30(1): 1265–1300. <https://doi.org/10.1007/s10639-024-12878-7>
- [33] Pellas N, 2023, The Effects of Generative AI Platforms on Undergraduates' Narrative Intelligence and Writing Self-Efficacy. *Education Sciences*, 13(11): Article 1155. <https://doi.org/10.3390/educsci13111155>
- [34] Song C, Song Y, 2023, Enhancing Academic Writing Skills and Motivation: Assessing the Efficacy of ChatGPT in AI-Assisted Language Learning for EFL Students. *Frontiers in Psychology*, 14, Article 1260843. <https://doi.org/10.3389/fpsyg.2023.1260843>
- [35] Liang J, Wang L, Luo J, et al., 2023, The Relationship Between Student Interaction with Generative Artificial Intelligence and Learning Achievement: Serial Mediating Roles of Self-Efficacy and Cognitive Engagement. *Frontiers in Psychology*, 14, Article 1285392. <https://doi.org/10.3389/fpsyg.2023.1285392>
- [36] Zhang Z, Wang C, Shang Y, 2026, Are AI Tools Effective in Reducing Cognitive Load or Boosting Writing Self-Efficacy for L2 Postgraduates? *The Asia-Pacific Education Researcher*, 35(2): 343–354. <https://doi.org/10.1007/s40299-025-01031-8>
- [37] Lin CY, Wang CC, 2025, Why Do Graduate Students Use Generative AI in Thesis Writing? The Influence of Self-Efficacy, Time Pressure, and Trust. *Current Psychology*, 44(12): 12071–12086. <https://doi.org/10.1007/s12144-025-08003-7>
- [38] Zhang L, Xu J, 2025, The Paradox of Self-Efficacy and Technological Dependence: Unraveling Generative AI's Impact on University Students' Task Completion. *The Internet and Higher Education*, 65, Article 100978. <https://doi.org/10.1016/j.iheduc.2024.100978>
- [39] Qin J, Karabacak E, 2010, The Analysis of Toulmin Elements in Chinese EFL University Argumentative Writing.

- System, 38(3): 444–456. <https://doi.org/10.1016/j.system.2010.06.012>
- [40] Toulmin S, 2003, *The Uses of Argument* (Updated ed., Cambridge University Press.
- [41] Hyland K, 2005, *Metadiscourse: Exploring Interaction in Writing* (1st ed., Continuum.
- [42] Pessoa S, 2017, How SFL and Explicit Language Instruction Can Enhance the Teaching of Argumentation in the Disciplines. *Journal of Second Language Writing*, 36, 77–78. <https://doi.org/10.1016/j.jslw.2017.05.004>
- [43] Jin T, Su Y, Lei J, 2020, Exploring the Blended Learning Design for Argumentative Writing. *Language Learning Technology*, 24(2): 23–34. <https://doi.org/10.64152/10125/44720>
- [44] van Merriënboer JJG, Kirschner PA, Kester L, 2003, Taking the Load Off a Learner's Mind: Instructional Design for Complex Learning. *Educational Psychologist*, 38(1): 5–13. https://doi.org/10.1207/S15326985EP3801_2
- [45] Guo Y, Lee D, 2023, Leveraging ChatGPT for Enhancing Critical Thinking Skills. *Journal of Chemical Education*, 100(12): 4876–4883. <https://doi.org/10.1021/acs.jchemed.3c00505>
- [46] Zhang R, Zou D, Cheng G, 2024, Chatbot-Based Learning Of Logical Fallacies in EFL Writing: Perceived Effectiveness in Improving Target Knowledge and Learner Motivation. *Interactive Learning Environments*, 32(9): 5552–5569. <https://doi.org/10.1080/10494820.2023.2220374>
- [47] Minh AN, 2024, Leveraging ChatGPT for Enhancing English Writing Skills and Critical Thinking in University Freshmen. *Journal of Knowledge Learning and Science Technology*, 3(2): 51–62. <https://doi.org/10.60087/jklst.vol3.n2.p62>
- [48] Hikmah D, Walida B, 2024, The Role of ChatGPT in Enhancing Students' Critical Thinking in Academic Writing. *Ethical Lingua: Journal of Language Teaching and Literature*, 11(2): 554–564. <https://doi.org/10.30605/25409190.780>
- [49] Emran AQM, Aldallal A, Nadheer A, 2024, Investigating the Impact of ChatGPT on Enhancing University Students' Critical Thinking Skills. In A. Harraf A. Hamdan (Eds.): *Business Development via AI and Digitalization* (Vol. 538, pp. 567–574, Springer. https://doi.org/10.1007/978-3-031-62102-4_47
- [50] Guo K, Zhong Y, Li D, et al., 2023, Effects of Chatbot-Assisted In-Class Debates on Students' Argumentation Skills and Task Motivation. *Computers Education*, 203, Article 104862. <https://doi.org/10.1016/j.compedu.2023.104862>
- [51] Liang W, Wu Y, 2024, Exploring the Use of ChatGPT to Foster EFL Learners' Critical Thinking Skills from a Post-Humanist Perspective. *Thinking Skills and Creativity*, 54, Article 101645. <https://doi.org/10.1016/j.tsc.2024.101645>
- [52] Zhang Q, Siraj SB, Abdul Razak RB, 2025, Effects of AI Chatbots on EFL Students' Critical Thinking Skills and Intrinsic Motivation in Argumentative Writing. *Innovation in Language Learning and Teaching*, 1–29. <https://doi.org/10.1080/17501229.2025.2515111>
- [53] Crossley SA, 2020, Linguistic Features in Writing Quality and Development: An Overview. *Journal of Writing Research*, 11(3): 415–443. <https://doi.org/10.17239/jowr-2020.11.03.01>
- [54] Nation ISP, 2022, *Learning Vocabulary in Another Language*. Cambridge University Press. <https://doi.org/10.1017/9781009093873>
- [55] Albrechtsen D, Haastrup K, Henriksen B, et al., 2008, *Vocabulary and Writing in a First and Second Language: Processes and Development* (1st ed., Palgrave Macmillan. <https://doi.org/10.1057/9780230593404>
- [56] Milton J, Wade J, Hopkins N, 2018, Aural Word Recognition and Oral Competence in English as a Foreign Language. In C. Abello-Contesse, R. Chacón-Beltrán, M. del M. Torreblanca-López (Eds.): *Insights into Non-Native Vocabulary Teaching and Learning* (pp. 83–98, Multilingual Matters. <https://doi.org/10.1017/9781009093873>

org/10.21832/9781847690290-007

- [57] Nation ISP, 2013, *Learning Vocabulary in Another Language* (2nd ed., Cambridge University Press. <https://doi.org/10.1017/CBO9781139524759>
- [58] Staehr LS, 2008, Vocabulary Size and the Skills of Listening, Reading and Writing. *Language Learning Journal*, 36(2): 139–152. <https://doi.org/10.1080/09571730802389975>
- [59] Laufer B, Nation P, 1999, A Vocabulary-Size Test of Controlled Productive Ability. *Language Testing*, 16(1): 33–51. <https://doi.org/10.1177/026553229901600103>
- [60] Cobb T, 2000, One Size Fits All? Francophone Learners and English Vocabulary Tests. *Canadian Modern Language Review*, 57(2): 295–316. <https://doi.org/10.3138/cmlr.57.2.295>
- [61] 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, Article 102221. <https://doi.org/10.1016/j.system.2020.102221>
- [62] Jacobs HL, 1981, *Testing ESL Composition: A Practical Approach*. Newbury House Publishers.
- [63] Gayed JM, Carlon MKJ, Oriola AM, Cross, J. S, 2022, Exploring an AI-Based Writing Assistant's Impact on English Language Learners. *Computers and Education: Artificial Intelligence*, 3, Article 100055. <https://doi.org/10.1016/j.caeai.2022.100055>
- [64] Nilson LB, 2013, *Creating Self-Regulated Learners: Strategies to Strengthen Students' Self-Awareness and Learning Skills*. Routledge. <https://doi.org/10.4324/9781003443803>

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.