International Journal of Computer-Assisted Language Learning and Teaching
Volume 14 • Issue 1 • January-December 2024

ChatGPT vs Teacher Roles in Developing EFL Writing

Sirin Sawangwan
https://orcid.org/0009-0009-5927-9046
Rajamangala University of Technology Tawan-Ok, Thailand

ABSTRACT

This study statistically compared the differences between Chat Generative Pre-Trained Transformer (ChatGPT) and teacher instruction, comparing the use of ChatGPT with teachers' roles in developing English as a Foreign Language writing. A mixed-method approach was employed. The quantitative component examined learners' writing performance, before and after revisions given by ChatGPT, as well as before and after receiving teacher instruction. Additionally, the study explores the differences between the results in the two groups. The qualitative analysis explored outcomes of using ChatGPT and identified teachers' roles through observational data. A total of 50 English-language learners from a public university were involved as a sample. The results, based on Test of English as a Foreign Language Internet-Based Test (TOEFL-iBT) criteria, revealed that teacher instruction led to higher scores and greater improvement than the use of ChatGPT revisions. Key implications of this study highlight teachers' roles in areas ChatGPT cannot replicate: (a) understanding writing criteria, (b) developing individual critical thinking, and (c) offering ethical guidance. Integrating ChatGPT into writing curricula and AI professional development are recommended.

KEYWORDS

ChatGPT, CALL, AI for Language Writing, Teachers' Roles, EFL Writing

CHATGPT VS TEACHER'S ROLES IN DEVELOPING EFL WRITING

In terms of English proficiency, many countries that do not have English as their dominant language are ranked low to very low in English language proficiency in the English as a second and foreign language membership, according to the English Proficiency Index (Education First, 2023). In other words, such countries will fall behind on the world economic stage if English teachers and learners cannot improve their use of English as a tool for advancement through education.

Fortunately, incorporating AI into English as a Second Language (EFL) curricula opens an opportunity for learners to improve their English proficiency. Currently, EFL educators are grappling with how to integrate AI into the development of English language skills, whilst maintaining academic integrity and addressing concerns such as critical thinking and personalized instruction.

The rapid transition to learning with AI means that teachers are challenged with new teaching and learning methods, and assessment protocol. An online platform powered by a large AI language model like Chat Generative Pre-Trained Transformer (ChatGPT) plays a vital role in EFL writing development, providing valuable opportunities for EFL learners to enhance their English proficiency (Masoudi, 2024; Meniado et al., 2024).

DOI: 10.4018/IJCALLT.361235

However, due to concerns regarding the use and misuse of ChatGPT, education institutions in many countries have banned the use of the chatbot (Reuters, 2023). For example, Alharbi (2024) addressed the gap between students and teachers perceiving different awareness and ethical concerns in using AI. Additionally, Tlili et al. (2023) reported issues related to cheating, honesty and truthfulness, misleading privacy, and manipulation with respect to the use of ChatGPT. Teachers and administrators see ChatGPT as opening a Pandora's box, posing threats to learners' development of critical thinking and writing abilities (Hong, 2023).

The potential for the use and misuse of AI encourages language educators to develop greater awareness, through questioning the use of AI in teaching languages. Are teachers still necessary, especially in the digital age? What specific roles should educators play in fostering learners' English writing skills, alongside technological advancements? Language teachers need to accept that they cannot fully control the use of AI in learners' English writing. The challenge, then, is how teachers can use AI as a good facilitator in assessing learners' writing skills, while also balancing ethical considerations, cultivating learners' critical thinking skills, and facilitating authentic writing—all of which are best taught by teachers.

To facilitate the development of writing competency among EFL learners, it is important to first clarify the potential roles of teachers, in comparison with ChatGPT. The following study provides a direct comparison between ChatGPT and teachers to determine which is the more effective in improving EFL writing skills.

Research Objectives

The aim of this study was to statistically compare the differences between ChatGPT and teacher instruction, and to compare the outcomes of using ChatGPT with teachers' roles in developing EFL writing skills.

Research Questions

In this study, the author sought to answer three research questions. First: Are there any significant differences in the overall English writing performance of learners before and after using ChatGPT, as well as before and after receiving teacher instruction? Second: Is there a statistically significant difference in overall English writing performance between learners who used ChatGPT for revision and those who made revisions under direct teacher instruction? Third: What are the outcomes documented in observational records that compare the roles of ChatGPT with that of teachers?

LITERATURE REVIEW

English Language Writing

Flower and Hayes (1980), through their protocol analysis, identified three components of the writing process: (a) the writer's long-term memory, (b) the writing processor, and (c) the task environment. The writer's long-term memory consists of knowledge of the topic, audiences, and writing plans. The writing processor has three components: planning, translating, and reviewing. The last operational process, the task environment, comprises two components: the rhetorical problems, and the text produced. In this study, we examined the role of the writing processor, investigating whether ChatGPT or a teacher serves as a more effective facilitator in developing English writing skills in EFL contexts.

The first key component, the writer's long-term memory, is crucial in helping EFL learners develop writing skills. Knowledge of an essay topic falls into this category. According to the Teaching English to Speakers of Other Languages International Association (2024), ChatGPT can generate topics for discussion and creative writing prompts, so that learners can improve their ability to write creatively. In contrast, teachers might be the best facilitators with respect to planning and providing

direct feedback. Additionally, teachers are better placed to discuss ethics in writing, such as issues concerning copyright and plagiarism (Hong, 2021). It is assumed that helping learners' long-term memory using both ChatGPT and teacher instruction can help students develop their English writing skills. Rudolph et al. (2023) proposed that learners can benefit from experiential learning through ChatGPT, which can generate various problem-solving scenarios and provide personalized tutoring. Skehan (1986, cited in Bigelow & Watson, 2013) asserted that teachers can help build confidence in language use and develop student aptitude. All these studies indicate that, overall, problem-solving enables learners to construct a language product—that is, writing.

Swain (1995) stated that in developing second and foreign language output, the first goal is to focus on the process of acquiring language, rather than only the product. The writing process stage—planning, translating, and reviewing EFL writing—falls into a similar category. The focus of the following study was whether ChatGPT or a teacher would be the better processor in helping learners plan, translate, and review their English writing.

From Computer-Assisted Language Learning to ChatGPT

This next section provides a concise, historical overview of the evolution from Computer-Assisted Language Learning (CALL) to advanced AI technologies like ChatGPT. It also reviews relevant studies on the advantages and limitations of ChatGPT, with particular emphasis on the central focus of this research.

CALL is defined as the search for, and study of, computer applications in the teaching and learning of language. It is the most important and relevant means of using technology in language learning today. CALL's premise is that technology is there to serve language learning, not vice versa (Levy, 1997, 2009). CALL was developed in the 1950s and continues to be developed in the 21st century. According to Warschauer (1996), CALL now encompasses four phases: behavioristic, communicative, integrative multimedia, and integrative internet. Today, CALL is no longer a discipline restricted to digital technology in formal language education, but has been adopted into mobile applications applied in learning and daily life (Chen et al., 2021). Language learners now enter a world in which their language literacy depends on CALL technology.

ChatGPT offers itself as a state-of-art technology for education. When users input a prompt into ChatGPT, it analyzes the input and rapidly generates a response drawn from information acquired, via machine learning in engagement with the internet ("ChatGPT," 2024). Current versions (3.5 and 4.0) have been pre-trained on massive amounts of data, allowing them, when prompted, to learn language patterns and associations used to generate human-like interactive text. ChatGPT excels at processing, distilling, and presenting information verbally, in human-like text. It is, therefore, an important asset for academic writers, particularly when users must outsource some of their written works (Pavilik, 2023, cited in Lingard, 2023).

Because ChatGPT works on a machine-based learning system, it has benefits and constraints. Lingard (2023) summarized its capacity and limitations: "ChatGPT generates responses that are grammatically correct and semantically meaningful. It can cultivate a piece of writing such as brainstorming ideas, generate counterarguments, summaries, and abstracts, edit to improve clarity and coherence" (pp. 261–270). ChatGPT can generate coherent and fluent text on various topics (Baskara & Mukarto, 2023), but the free version (3.5) can produce responses of only 500 or fewer words. Furthermore, it has limitations with respect to integrity and plagiarism (Perkins, 2023).

Examining the history of CALL, in relation to AI tools like ChatGPT, reveals that the use of technology tends to shorten the process of English language development. According to Crystal (2004), writing with technology lessens the length of English language development, regardless of how far that technology can be developed. Innovation is then anticipated in each of the three traditional domains of language and communication: production, transmission, and reception. Consequently, a key goal of the research in this article is to critically assess which of these areas ChatGPT might respond to most effectively.

Effectiveness of ChatGPT in English Writing

This section illustrates how ChatGPT can enhance development of language learning skills, particularly in the context of EFL. By providing automated assistance in language acquisition, ChatGPT offers opportunities for learners to practice and enhance their linguistic abilities, particularly in writing. Several notable studies are reviewed to highlight the potential role of ChatGPT in supporting EFL learners' development of English writing skills. These studies serve to introduce the broader possibilities of integrating artificial intelligence into language instruction, positioning ChatGPT as a supplementary tool for fostering learner autonomy and enhancing writing proficiency.

In 1978, Vygotsky's sociocultural concept of scaffolding theory was shown to help learners develop cognitive intelligence. He believed that learning occurs at two levels: individual and social with effective helping aids. Attention to both, these levels develops the Zone of Proximal Development (ZPD) in learners. Without the ZPD, students might fail to increase their learning capability. ChatGPT can facilitate the advancement of the ZPD by providing rapid feedback, when compared with that available through teacher instruction—which requires more time in which to build up the ZPD. The study presented here explores whether ChatGPT or a teacher can be a more effective facilitator of student learning.

ChatGPT can be used in developing learners' language skills, to scaffold the learning process by providing feedback to students on their language use, and acting as a partner in practicing language, by recommending activities for more language practice (Bin-Hady et al., 2023). ChatGPT can generate writing prompts and provide feedback on written work, both of which can improve learners' writing skills (Baidoo-Anu & Owusu Ansah, 2023). Moreover, it considers the event and writing order, including the use of main findings, explanatory sentences and a conclusion (Rudolph et al., 2023). It also considers verb tense, as related to the given topic (Fitria, 2023). Moreover, Baskara and Mukarto (2023) assert that ChatGPT improves both the efficacy of and engagement in language instruction.

Numerous research studies have explored the effects of ChatGPT on English language learning outcomes. Li et al. (2023) found that all the participants who used ChatGPT demonstrated different degrees of improvement in their writing scores. Song and Song (2023) compared writing papers between participants in both control and experimental groups. The results showed that the overall English writing pre-test scores of the experimental group increased from 39.26 to 59.12. In addition, Huang et al. (2023) indicated that AI was frequently used to assist students in learning writing, reading, vocabulary, grammar, speaking, and listening. as compared with their counterparts in a non-AI course. Furthermore, Yan (2023) revealed that a potential applicability of ChatGPT is in developing a second, language writing pedagogy.

The outcomes of these studies are significant, for both EFL educators and learners. Notably, the use of ChatGPT in an EFL classroom has yielded compelling evidence in favor of English writing proficiency. Furthermore, by endorsing collaborative writing facilitated by ChatGPT, educators can cultivate and sustain learners' enthusiasm and interest in writing outcomes, and ultimately scaffold students' overall growth and writing proficiency.

Teachers' Roles in Technology Use

There have been controversies inherent in the use of technologies in EFL writing. Should teachers be replaced with AI in terms of teaching methods and assessments, for example? How should AI be appropriately used in writing? To begin to answer these questions, teachers must be competent in the use of the relevant technology. Accordingly, the International Society for Technology in Education (ISTE) standards for educators categorize what competent teachers should know and should be able to do with technological competencies (ISTE, 2024). These standards are categorized under the following headings:

- Learner

- Leader
- Citizen
- Collaborator
- Designer
- Facilitator
- Analyst

Research studies have emphasized these seven key areas as listed above. Studies have shown that teachers and learners seek out and adapt to new technologies. Regarding learners, collaborators, and facilitators, Chiu et al. (2023) revealed that intrinsic competence in learning with a chatbot depended on both teacher support and student expertise (i.e., self-regulated learning and digital literacy). Additionally, teacher support better satisfied the need for relatedness, and offered less satisfaction regarding the need for autonomy. Wang and Kruk (2024) revealed a strong link between teacher confirmation and credibility and teacher-to-student academic engagement. Apridayani et al. (2024) suggested that problem-solving, a fundamental skill in writing, should utilize a combination of teacher and student assessment, and doing so affected the good quality of student essays.

Teachers should encourage learners to participate responsibly in the digital world. Perkins (2023) found out that online assessment was associated with increased risk of compromising academic integrity. Studies have indicated that learning with technological environments and classifiers are able to accurately detect whether paragraphs were written by ChatGPT or a human, findings that are important for curriculum designers to consider (Berriche & Larabi-Marie-Sainte, 2024).

Analysts, too, benefit from the work of St-Onge et al. (2022), who proposed four cautious, and technology analytical concerns about how to effectively use data for e-assessment in higher education: (a) consequences for how learners decide to move forward, (b) pre-occupation with potential cheating, (c) the importance of pedagogical alignment, and (d) the affordances available to them.

Although meeting the ISTE standards for teachers' technology competencies is important, the key role of a teacher in the digital age is to go beyond what AI cannot accomplish. One primary focus of this study, therefore, was to explore how teachers might foster the ISTE standards—and go even further.

METHODS

The research documented in this article utilized a mixed methodology. In response to the first research objective, quantitative methods were used to statistically compare the differences in EFL learners' writing development, when utilizing ChatGPT versus following a teacher's instruction. Meanwhile, in attending to the second research objective, a qualitative method was used to discern the difference between utilizing ChatGPT, as compared with the role of teacher in developing EFL writing.

Data Collection

Data collection was divided into two parts: (a) the population under study and (b) participant selection criteria.

Population Under Study

The population of this study comprised EFL undergraduate learners enrolled in an English language course at a public university. Before enrolling in these courses, the learners had to pass a Fundamental English compulsory course, which is required to be completed during either the first or second semester of their first year. The compulsory examination used for the midterm and final examinations of Fundamental English is designed in compliance with the course syllabus, which also includes writing. In addition, all learners who pass the writing section take the same standard

examinations for both the midterm and final examinations. Only those learners who gained a score of 17.5, 18, or 18.5 out of 25 on the writing part of both exams were included in the population of this study. Using these conditions, it was possible to assume that these learners possessed a similar competence level in English writing.

Participant Selection Criteria

The researcher applied a purposive sampling method to recruit 50 learners for this study; this number of participants was chosen due to time constraints and manageability issues.

In the first semester of the 2024 academic year, the total number of the learners enrolled in English courses was approximately 1,000. According to the section system currently available, these learners were divided into 40 sections, and each section had approximately 25 learners. The researchers identified sections in which were enrolled the most participants who scored between 17.5 and 18.5 on the compulsory writing exams; five sections out of the 40 were then selected. The 50 learners were divided into two groups: the ChatGPT group, and the teachers' instruction group. All learners were volunteers who gave consent to the ethical considerations of this study and felt comfortable using computer-based tests.

Instruments

The Internet-Based Test

The Educational Testing Service (2023) standards indicate that writing performance can be measured by writing a 300-word independent essay in 30 minutes Accordingly, participants were assigned a 300 word-essay for three tests: (a) a pre-test, (b) revision from ChatGPT, and (c) revision under teacher instruction on the same topic: "What is your favorite subject, and why?"

A Test of English as a Foreign Language Internet-Based Test (TOEFL-iBT) Independent Writing Rubrics measure (2023) was used to determine participants' essays scores on all tests, with raw scores presented in Figure 1.

After the resulting scores were obtained, the raw scores were converted into scaled scores (see Figure 3). Finally, the researchers compared those scores with the English proficiency skill according to the Common European Framework of Reference for Languages (CEFR); these are shown in Figure 4.

With regards to scoring validity, researchers summarized and used three marking criteria from the Educational Testing Service (2023). These were: development, organization, and language use.

For reliability, the author requested three inter-raters who were experts in testing, as well as in teaching English writing skills. These experts graded the pre-tested participants, the revisions from ChatGPT, and the revisions from teacher instruction.

ChatGPT

The researchers used ChatGPT, organized by OpenAI (Version 3.5), in the revision, with the following prompt: "Revise this essay to 300 words, adhering to TOEFL iBT criteria: development, organization, and language use", followed by each learner's pre-test essay. IBM SPSS Statistics v26 was used to compare mean scores for all tests.

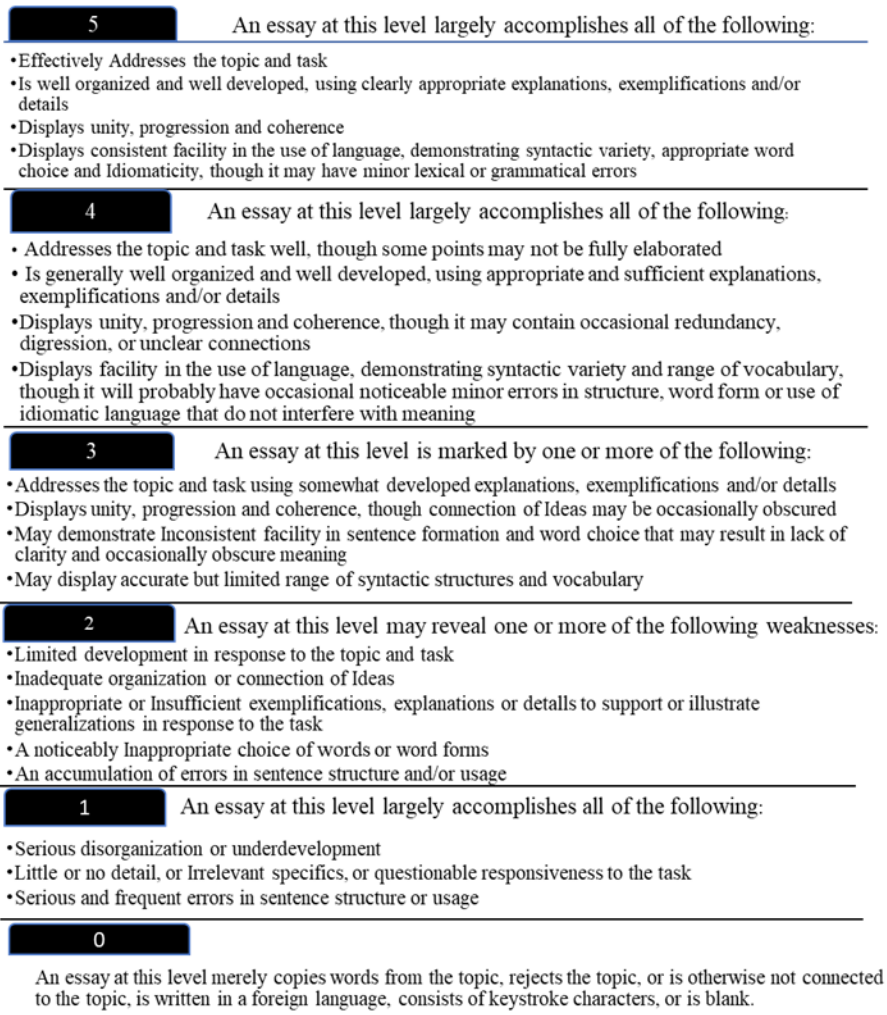
Observation Records

Observation records were obtained from field notes, and then categorized into five reports: (1) a comparative analysis of the percentage of writing errors, (2) examples of writing development, (3) reports of the misuse of gender pronouns, (4) reports of technical issues, and (5) reports of ethical issues.

Data Collection Procedure

The data collection procedure is outlined in Table 1.

Figure 1. Writing raw score *rubric*



Note. Adapted from Educational Testing Service (2023). TOEFL Criteria Marking Scheme. https://leapscholar.com/blog/toefl-writing-score-2/

In Steps 1 and 2, participants were asked to write, using Microsoft Word, a 300-word essay in 30 minutes, on the topic of "What is your favorite subject, and why?" Essays were then given to the three inter-raters, who scored them using Writing Raw Score Rubrics (see Figure 1). The inter-raters returned their marked scores, comments, and feedback to the researcher.

In Step 3, revisions were assigned to the first group after they finished their pre-test, using ChatGPT. These participants were provided with a prompt that read: "Revise this essay in 300 words, adhering to TOEFL iBT criteria: development, organization, and language use." In this step, field notes were taken to record issues that occurred while participants used ChatGPT.

In Steps 4 and 5, revised essays using ChatGPT were again sent to the inter-raters to score. The inter-raters returned the marked essays, with comments and feedback, to the research team, who then compared the mean scores of the pre-test and the ChatGPT revision, using IBM SPSS Statistics v26.

In Step 6, all comments and feedback for the pre-test and the ChatGPT revisions from the three inter-raters were collected and read by the research team. They then summarized the feedback using

Table 1. Data collection procedure

Step	Procedure
1	Pre-test: 50 learners wrote a 300-word essay on "What is your favorite subject, and why?"
2	Three inter-raters marked participants' writing scores with comments and feedback.
3	The first group revised their essays by using ChatGPT (field notes recorded).
4	The three inter-raters marked participants' scores with comments and feedback.
5	Mean scores compared between the pre-test and ChatGPT.
6	Returned all comments and feedback to the second group.
7	The second group revised their essays according to teacher instruction (field notes recorded).
8	The three inter-raters marked participants' scores with comments and feedback.
9	Mean scores compared between the pre-test and revision under teacher instruction.
10	Two groups' mean scores compared.

keywords, highlighted errors, sentence corrections, and all comments from the inter-raters; these were then returned these to the second group.

In Step 7–9, revisions were administered to the second group, under teacher instruction, by having the participants revise their pre-test essays, according to feedback and comments provided. Their English writing teachers then trained students on how to revise their essays, according to the comments and feedback. The revised papers were then returned to the inter-raters to score. The inter-raters marked the participants' papers and returned the scores, with comments and feedback, to the researchers—who also took field notes. The mean scores of the pre-test and the revised essays were compared, according to the teachers' advice.

Once again, IBM SPSS Statistics v26 was used to compare mean scores between the two groups.

Quantitative Analysis

In response to the first research question, paired-samples *t*-test using IBM SPSS Statistics v26 were conducted to compare the mean scores of the pre-test and the ChatGPT revisions, as well as between the pre-test and the teacher instruction revisions. For the second research question, an independent-samples *t*-test was conducted, to compare the mean scores between the two groups.

Qualitative Analysis

To answer the third research question, observational records were reported from field notes. Merriam (2009) describes a four-step process of field note analysis: (a) reviewing, (b) analyzing, (c) categorizing, and (d) reporting. At the first step, researchers read through every marked score and errors on the pre-test, and then separated the essays into two groups: revision from ChatGPT, and revision from teacher instruction. Next, they analyzed the percentage of writing errors among participants in the two groups, based on the marking and the scoring criteria (development, organization, and language use). Analyzed errors and other issues found throughout the experiment were categorized. Finally, researchers compared the percentage of the errors on the pre-test between the two groups. This was done through a comparative report, while other issues and notable essay extracts were summarized. These findings are described in the Qualitative Findings section.

Table 2. Results of pre-test and revisions from the ChatGPT group

Group	*N*	*M*	*SD*
Pre-test	25	2.46	0.54
Revision	25	3.92	0.96*

Note. N = number; M = mean, SD=Standard Deviation.

Table 3. Paired-samples t-test results for pre-test and revision of the ChatGPT group

ChatGPT	Paired differences				*t*	*df*	*Sig.* (2-tailed)
	M	*SD*	*SEM*	95% CI of the difference			
Pre-test revision	1.46	0.62	0.12		–11.78	24	.000*
				[–1.71, –1.20]			

Note. M = mean; SEM = standard error mean; CI = confidence interval.
*$p < .001$.

Table 4. Result of pre-test and revision from teachers' instruction scores of the teachers' instruction group

Group	*M*	*SD*
Pre-test	2.57	0.40
Revision	4.87	0.26

Note. M = mean.

QUANTITATIVE RESULTS

Paired-Samples and Independent-Samples *t*-Test Reports

Table 2 shows the mean writing performance scores of the ChatGPT group. The learners' mean pre-test score was 2.46, and the mean of the ChatGPT revision score was 3.92. The pre-test standard deviation (SD) was 0.54, and the ChatGPT revision SD was 0.96. These results show that there was an increase in the mean scores between pre-test and ChatGPT revision at 1.46.

As shown in Table 3, the *t*-test analysis indicates an increased mean score from pre-test (2.46) to ChatGPT revision (3.92), which is 1.46, which is statistically significant at the $p = .000$ level. The differences between the mean of the pre-test and ChatGPT revision scores were reported to be –1.46, $t(df24) = –11.78$, Sig. = .000, $p < .001$. This suggests that there is a highly significant difference in learners' overall English writing performance, both before and after ChatGPT revision.

Remark: * a deviated data is explained in the discussion section titled AI Professional Development.

Table 4 shows the mean scores of the writing performance of the teacher instruction group. The mean pre-test score was 2.57, and the mean teacher instruction score was 4.87. The pre-test SD was 0.40, and the revision SD was 0.26. These results show that there was an increase of 2.30 in the mean scores between pre-test and teacher instruction revision.

As seen in Table 5, the *t*-test analysis indicated that an increase in mean score from pre-test (2.57) to teacher instruction (4.87) was 2.30, which is statistically significant at the $p = .000$ level. The difference between the mean pre-test and teacher instruction scores was –2.30, $t(24) = –24.72$, Sig. = .000, $p < .001$. This can be interpreted as indicating that there was a highly significant difference in learners' overall English writing performance, both before and after receiving teacher guidance.

Table 6 shows the mean writing performance scores of both groups. The ChatGPT mean score was 3.92, whereas the mean of the teacher's instruction score was 4.87. Standard deviations were 0.96 and 0.26, respectively. These results suggest that the mean English writing performance scores

Table 5. Paired-samples t-test results for pre-test and teacher instruction revision group

Teacher instruction	Paired differences				*t*	*df*	*p* (2-tailed)
	M	*SD*	*SEM*	95% CI of the difference			
Pre-test revision	2.30	0.46	0.93		–24.72	24	.000*
				[–2.49, –2.10]			

Note. M = mean; SEM = standard error mean; CI = confidence interval, df = degree of freedom.
*p < .001.

Table 6. Mean scores of the ChatGPT and teacher instruction groups

Group	*N*	*M*	*SD*	*SEM*
ChatGPT	25	3.92	0.96*	0.19
Teacher's Instruction	25	4.87	0.26	0.05

Note. N = number; M = mean; SD = Standard Deviation, SEM = standard error mean.

Table 7. Mean scores of the ChatGPT and teacher instruction groups: Independent-samples t-test results

Score	Levene's test for equality of variances		*t*-Test for equality of means					
	F	*p*	*t*	*df*	*p*	Mean difference	Standard error difference	95% CI of the difference
Equal variances assumed	14.67	.000*	–4.73	48	.000*	–0.94	0.19	[–1.34. –0.54]
Equal variances not assumed			–4.73	27.52	.000*	–0.94	0.19	[–1.35. –0.53]

Note. CI = confidence interval.
**p* < .001.

of learners in the teacher instruction group were higher than those in the ChatGPT group, at mean score of 0.95.

Table 7 shows the results of an independent *t*-test of the revision scores of the two groups. The result F=14.67 $t(df48)$ = –4.73, sig..000, p= <.001) shows that there was a significant difference between the ChatGPT and the teacher instruction groups. This suggests that learners who received teacher instruction had significantly higher revision scores, as compared with those who used ChatGPT.

A Comparison Between the Two Groups' Development

Figure 2 illustrates that the mean difference of the pre-test (2.46) and ChatGPT revision (3.92) is 1.46, whereas the mean difference between the pre-test (2.57) and teacher instruction (4.87) is 2.30 out of 5. These results indicate that there is a greater improvement in writing after the teacher instruction than after ChatGPT revision.

TOEFL-iBT Scores

The pre-test results of both groups (2.46 and 2.57) fall into a TOEFL-iBT writing rubric mean of 2.50 out of 5, which can be converted to a scaled score (see Figure 3) of 17 out of 30. In contrast, after revision under teacher instruction, a raw score of 4.87 corresponds to a scaled score of 29. On the other hand, after ChatGPT revision, a raw score of 3.92 corresponds to a scaled score of 25.

Learners improved their English writing performance from a score of 17 to 29, reflecting an increase of 12 points, which corresponds to a 70.59% development after teacher instruction. Meanwhile, after ChatGPT revision, their writing performance increased by 8 points, or 47.05%.

Figure 2. A comparison between the pre-test and the revision mean scores of the two groups: Overall English writing performance

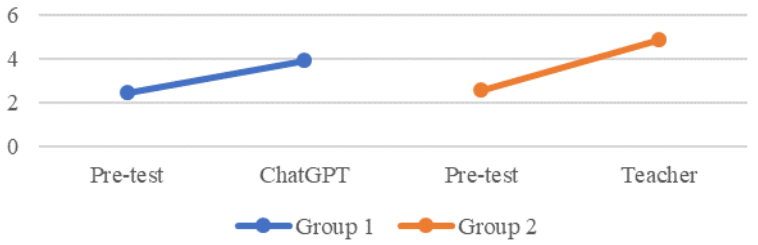
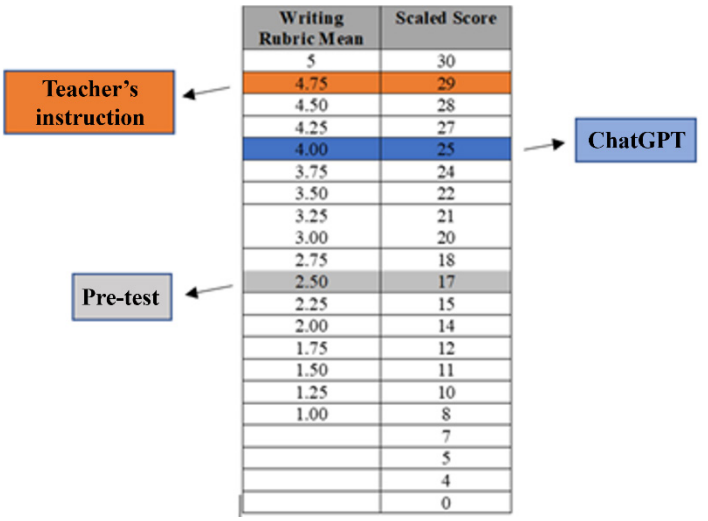


Figure 3. TOEFL-iBT scaled scores rubric



Writing Rubric Mean	Scaled Score
5	30
4.75	29
4.50	28
4.25	27
4.00	25
3.75	24
3.50	22
3.25	21
3.00	20
2.75	18
2.50	17
2.25	15
2.00	14
1.75	12
1.50	11
1.25	10
1.00	8
	7
	5
	4
	0

Note. Adapted from LeapScholar (2024). TOEFL writing score: Latest scoring system & marking scheme of 2023 explained. https://leapscholar.com/blog/toefl-writing-score-2/

It can be concluded that the overall English writing performance of EFL learners improved more statistically and substantially after they had received instruction from a teacher, compared with using ChatGPT for revision.

A Shift in English Language Writing Competency

To compare TOEFL-iBT with CEFR levels, the pretest result (17) was multiplied with the four English language skills (listening, speaking, reading, and writing). The average calculation yielded a TOEFL-iBT score of 68, corresponding to a B1 level on the CEFR scale (see Figure 4).

The writing results after the use of ChatGPT and the teacher instruction with the combination of the four skills can be scored at 100 and 116, respectively. These conversion scores from both groups correspond to the C1 level on the CEFR scale.

One can conclude that the quantitative results demonstrate that learners' English writing performance improved from 68 (Intermediate) to 100 and 116 (Advanced), after the integration of ChatGPT and teacher instruction, respectively. This reflects a similar development level through the combined use of ChatGPT and teacher instruction, moving from the "Intermediate" (B1) level to the "Advanced" (C1) level.

Figure 4. A comparison between CEFR level with TOEFL-iBT

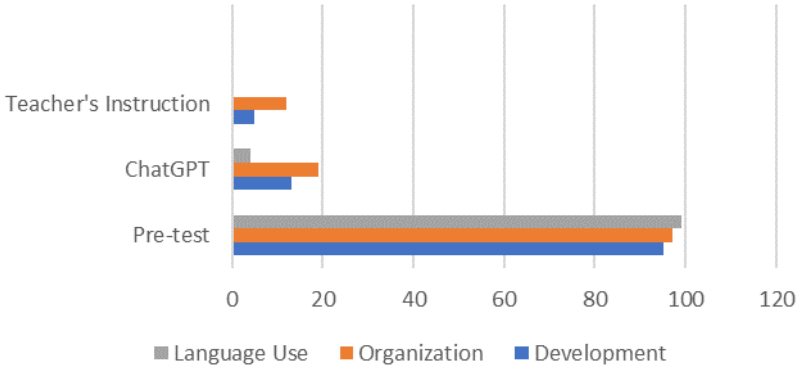
CEFR	TOEFL-iBT
Pre – A1	n/a
A1 Beginner	n/a
A2 Elementary	n/a
B1 Intermediate	42-71
B2 Upper Intermediate	72-94
C1 Advanced	95-120
C2 Proficient	n/a

Pre-test

Teacher's & ChatGPT

Note. Adapted from English First (2024). Converting from one English level to another. https://www.efset.org/english-score/#converting-from-one-english-level-to-another

Figure 5. Comparison between writing errors



QUALITATIVE FINDINGS

The qualitative outcomes drawn from the observational record (field notes) are presented in five reports: a comparative analysis of the percentage of writing errors, examples of writing development, reports of the misuse of gender pronouns, reports of technical issues, and reports of ethical issues.

Comparative Analysis of the Percentage of Writing Errors

Figure 5 presents a comparative analysis of the percentage of writing errors derived from the pre-test scores of all 50 participants, along with those in the ChatGPT revision group and the teacher instruction revision group.

As delineated in Figure 5, the examination of errors obtained from the returned comments and feedback revealed that participants in the pre-test phase exhibited ambiguous development at a rate of 95%, alongside deficiencies in organizational structure at a rate of 97%. Notably, most deductions during the pre-test were attributed to issues in language use, accounting for 99% of observed shortcomings.

For the first group, after ChatGPT revision, there was substantial decrease in comments and feedback pertaining to development and organization, plummeting to figures of 13% and 19%, respectively. In contrast, errors concerning language use diminished—a mere 4%.

After teacher instruction, the degree of comments and feedback regarding errors in development and organization reduced significantly, to 5% and 12%, respectively; errors related to language use were entirely eradicated.

Examples of Writing Development

Three notable examples demonstrating pre-test and revisions from ChatGPT group are presented in Extracts A and B. The Teacher's instruction group is elaborated in Extract C.

Extract A

This pre-test was scored as having not adhered to the topic, and demonstrating deficiencies in idea development, organization, and language use. It was given a score of 0.

Hi!, i'm … you can call me … I am … years old. I live in … But families in the provinces. My favorite color is … I like to be outside …

Although ChatGPT was used to help rewrite this essay, the required subject was still not addressed. Moreover, the word count was still too low after the revision, thus, grade deductions were made regarding essay development and organization. The use of ChatGPT resulted in an improvement in the language use, however, leading to a score of 1.

Hello I'm … but you can call me … I am … years old. I live in … But family resides in the provinces. Purple like lavender is my favorite color. I enjoyed spending time outdoor with friends.

Extract B

The pre-test in Extract B was marked by deficiencies in development. A lack of coherence in connecting ideas led to inadequate organization. Furthermore, grammatical errors were prevalent throughout the essay. It was given a score of 1.92.

I like this subject because the teacher is fun … The teacher is very kind. Teachers make learning English not boring. Can actually use what you study in everyday life.

After the ChatGPT revision, the key answer to the subject was still omitted and the demonstrative pronoun "this" was not eradicated. The word count was short by approximately fifty words, however, an improvement in the language use was noted, leading to a final score of 3.

I'm drawn to this subject because of the engaging teaching style of the instructor. The teacher consistently fosters an enjoyable … Additionally, this subject requires minimal homework, which aligns … Furthermore, this subject yields excellent results.

In Group 2, a participant's essay significantly improved from a score of 1.92 to 4.75, after the teacher's instructional guidance during the revision process.

Extract C

I like the field that I am studying in, which is … Because I learned both … Both difficult and easy words and can use those words again You can … Lastly, studying with the teacher was …

At first, this essay was marked as having not addressed the topic, with deficiencies in development and vague ideas—causing poor organization. After the teacher taught this participant how to write a good essay, they submitted the revision and received a score of 4.75.

I'm passionate about studying in the subject called Hotel Management. Combining both theory and hands-on experiences caught my attention. Despite enjoyable teaching environment, … Quick-paced lecture sometimes … Nonetheless, the relaxed atmosphere fosters enjoyment in learning the subject.

Reports of Misuse of Gender Pronouns

Two notable examples from the participants in ChatGPT group indicated ambiguity in pronoun usage.

Table 8. Reported technical issues with ChatGPT

Extracts	Technical issues with ChatGPT
A, B	Topic unidentified
A, B, F, G, H, I, J	Second prompt required for expanding the essays

Extract D

Before delving into your favorite subject …

Extract E

Due to teacher **Nancy**'s humorous and kind demeanor, **his** engaging teaching style.

Reports of Technical Issues

There were seven examples of technical issues during the ChatGPT revision, as elaborated in Table 8.

As indicated in Table 8, revisions to extracts A and B generated by ChatGPT were submitted without addressing a specific subject. According to the field notes, the given prompt clearly stated that the essay should be revised under topic's name. ChatGPT cannot provide its own answer responding to the topic; therefore, had the participants introduced an additional prompt, such as, "Relate this essay to the topic" along with "What is your favorite subject and why?'—Psychology and English, respectively," the inclusion of these subjects in the prompt may have enabled ChatGPT to enhance the quality of their essays.

For Extracts F–J, which are not presented here, the revisions generated by ChatGPT were submitted with low word counts, ranging between 50 and 100. We noted that if the participants had been provided a follow-up prompt, such as "Expand this essay to 350 words," the resulting technical issue could have been resolved, possibly improving the overall revision scores.

ChatGPT cannot cover all essay requirements. Learners still need to know how to overstate the prompt to respond to the essay topic and criteria.

Report of Ethical Issues

The revised introduction in Extract K, generated by ChatGPT, lacked proper citations. One can see that in case of citation, without guidance from the teacher, ChatGPT was unable to assist this learner in providing correct citations.

Extract K

Experts attribute EFL countries' low level of English proficiency to a combination of factors including structural problems in their education systems and cultural barriers.

In summary, ChatGPT is limited in its ability to provide instructions on essay writing criteria, identify gender pronouns, reorganize disorganized essays, and generate ethical citations—all areas that remain dependent on teacher guidance.

DISCUSSION

Identifying Teachers' Roles

The ISTE's (2024) seven standards for teachers in the digital age are no longer adequate. Teachers should go beyond what AI cannot replicate. Although significant improvements in writing performance have been observed using ChatGPT, and it can function as a rapid and effective personal language tutor, the teacher's role remains indispensable. The results of this study highlight the need

for teachers in the digital age to focus on areas where AI falls short—and cultivate the unique, human skills that technology cannot replicate. The research in this study found that ChatGPT could not teach students to fully comprehend writing criteria. Additionally, it was unable to facilitate the development of individual critical thinking. Finally, it could not provide ethical guidance in regarding academic integrity. These critical educational elements require the nuanced understanding and expertise that only human teachers can offer.

Understanding Writing Criteria

According to Flower and Hayes (1980), the writing process comprises three key components: planning, translating, and reviewing. While these elements are essential, they are not sufficient on their own. An understanding of writing criteria should be incorporated into the process. The study reveals that teachers, as more proficient processors, are better equipped to instruct learners on comprehending writing criteria, prior to the planning phase of writing.

Because ChatGPT is a machine learning-based form of AI, it can only offer responses based on information gleaned from existing, generalized data. It is incapable of helping users understand writing criteria, expressing emotions, making decisions (by stating what subject is favorable to the users), or providing subjective opinions, especially regarding personal preferences. The topic given for the tests in this study exemplifies these limitations, given that ChatGPT was unable to respond meaningfully to the essay prompt without being provided precise instructions. This strongly suggests that it is up to teachers to help learners understand writing criteria, before planning what and how to write.

According to the comparative analysis of the percentage of writing errors (Figure 5), there are three main criteria regarding how to teach students to improve their essay writing. These are development, organization, and language use.

Regarding development, learners who receive lower scores in this area should receive guidance to ensure that their essays consistently address the topic, beginning in the first paragraph. In addition, teachers should remind learners that an error in calculating the number of words can result in a loss of one to two points. During the drafting process, it is essential to calculate the number of words per sentence and the number of sentences per paragraph. Most importantly, learners should be trained to type 300 words in 30 minutes, covering all key points.

With respect to organization, given the persistent errors in organization shown in Figure 5, one can infer that ChatGPT is still unable to effectively reorganize a disorganized draft. Teachers should guide learners, therefore, in how to organize their essays by paragraphing and linking each paragraph to a single sub-idea. They should also encourage students to use ChatGPT to compare the revisions with the structure and organization of their original pre-test responses. In addition, teachers should provide instruction on how to align explanations and details with the main idea and sub-ideas. Guidance should also be given on the appropriate number of sentences needed to maintain well-organized paragraphs.

Regarding language use, and again referring to Figure 5, it is evident that ChatGPT significantly reduced most of the ambiguity in the students' responses. Teachers can further leverage the advantages of ChatGPT, by guiding learners in comparing their pre-test responses with ChatGPT-generated revisions, on both a word-by-word and sentence-by-sentence basis. This process will allow students to focus on their use of academic vocabulary, identifying grammatical errors, and assessing the appropriate use of idiomatic expressions.

Developing Critical Thinking

Swain (1995) asserted that a primary concern of second language acquisition is to develop effective output, but that we must consider the "'process,' NOT only the 'product'." Although revising with ChatGPT can produce a highly effective product", it cannot produce an effective *process* in regard to English writing practice—the process of writing can only be done well with teacher instruction.

In the development of critical thinking skills, only teachers can guide individual learners through the process of writing essays, step-by-step. Scaffolding, as an initial writing stage, helps students

generate a comprehensive response to any given topic (Vygotsky, 1978). From there, learners are taught to generate, and then organize, their main idea and sub-ideas, systematically building on them until they reach the final stage of formulating a conclusion. Although ChatGPT can contribute to the scaffolding process by generating a finished, good-quality paper as an example, it is unable to facilitate critical thinking, when the scaffold is removed and the learner's ZPD is allowed to grow independently. ChatGPT lacks the ability to foster deeper cognitive processes needed to generate original critical ideas, form rational arguments, and expand on sub-ideas. Only teachers can guide learners through this crucial developmental process. Ultimately, it is the teacher's influence that ensures critical thinking and the ability to write effectively.

Ethical Guidance

ChatGPT cannot provide ethical guidance, as shown from the research provided in this article. As per Lingard's assertion (2023), ChatGPT operates on a machine-based learning system, which offers both benefits and limitations. ChatGPT generates responses for grammatical and semantic purposes effectively but is limited in areas of integrity and plagiarism (Perkins, 2023). It cannot teach students how to prevent plagiarism, such as paraphrasing sentences while preserving the core meaning, and how to do in-text citations. Academic essays require thorough, accurate citations—and according to this study, teachers are still needed to guide learners regarding proper citation practices

AI Professional Development

Adopting new technology for education requires professional development. Teachers should be competent in using AI for their pedagogy. They should be not only instructors, but also able to offer technical support. Teachers need to understand the limitations of ChatGPT and know how to integrate ChatGPT into their courses; they should stay up to date with trends in using AI-assisted as an effective platform for EFL learners. Javaid et al. (2023) suggests that teachers can function as training assistants; Aithal and Aithal (2023) recommend that teachers in higher education should attend courses in data analytics, AI literacy, and digital literacy.

There exists significant variation in learners' abilities to effectively use ChatGPT. Table 8, for example, indicates that numerous instances were reported of students not reaching the required word count, despite very clear commands. This suggests that learners should be taught and trained to be aware of issues inherent in giving prompts for ChatGPT to create their essays. They should also learn more effective and complex ways to address AI prompts, such as "Revise this essay in 350 words," or "Write an essay on what is your favorite subject—English writing."

Implementing ChatGPT Into EFL Writing Curricula

Although teacher instruction facilitates EFL learners receiving improved TOEFL-iBT scores, it was noted from result (Figure 4) that a similarly tangible shift was achieved through use of ChatGPT, helping participants move from the Intermediate (B1) level to the Advanced (C1) level in CEFR. This supports Crystal's proposal that writing with technology shrinks the length of English language development (2004). From the findings (Figure 4), improvements in CALL for writing pedagogy correspond with higher levels in the language production domain; specifically, the test result indicates that ChatGPT can contribute to improved writing development in a short amount of time. Certainly, in this study, ChatGPT proved its potential in response to "product" question. Pursuant to the shift in improvement from B1 to C1 of CEFR levels, EFL universities should integrate ChatGPT into their English language writing curricula, under teacher supervision. By incorporating ChatGPT, the role of teachers as facilitators would evolve, focusing on providing writing guidance and personalized support.

The utilization of ChatGPT for initial assessments and feedback can save teachers valuable time normally spent on checking for errors. This time can then be redirected toward the planning and development of a more effective and engaging English writing curriculum, ultimately enhancing

Table 9. The superior roles of teachers compared with ChatGPT

ChatGPT	Teacher's roles
47.05% Better writing development	70.59% Better writing development
13% Errors in development	5% Errors in development
19% Errors in organization	12% Errors in organization
4% Errors in language use	0% Errors in language use
Unable to understand essay writing criteria	Able to provide instruction on essay writing criteria
Unable to make decisions or form opinions	Able to provide critical instruction
Misuse in grammar (pronouns)	Able to teach and correct grammar
Unable to reorganize a disorganized essay	Able to provide step-by-step essay writing instruction
Unable to provide ethical citations	Able to provide ethical instruction
Provide product of EFL writing	Provide process of EFL writing

learners' writing performance and overall language proficiency. The integration of such rapid technology with ChatGPT promises a more efficient and productive EFL educational environment.

English Proficiency and Economic Opportunities in EFL Countries

To avoid being ranked as "low" or "very low" in English proficiency, as indicated by the English Proficiency Index (Education First, 2023), countries may consider the findings of this study as offering valuable insights into the use of ChatGPT for improving EFL writing competency. However, when comparing the shift in skills from Intermediate to Advanced, it is evident that the use of ChatGPT is a more effective tool for enhancing the English proficiency of EFL undergraduate learners, when used under a teacher's guidance

CONCLUSION

The results of this study have demonstrated, quantitatively and qualitatively, that teachers contribute more to the development of EFL writing, when compared with ChatGPT. This conclusion is supported by the summarized data presented in Table 9.

The results of this study can be summarized in three main points: the overall English writing performance of EFL learners, the teacher's roles that ChatGPT cannot replicate, and the necessity of implementing ChatGPT into both EFL writing curricula and professional development for teachers in the realm of AI.

First, the overall English writing performance of EFL learners improved more statistically and substantially *after* they had received instruction from teachers, as compared with when they used ChatGPT for revision. This was according to TOEFL-iBT independent writing raw scores of 4.87 to 3.92, respectively. Moreover, greater improvement in writing essay raw scores was noted in the form of 12 points (70.59%) after teacher instructions, as compared with 8 points (47.05%) after revision of ChatGPT, according to TOEFL-iBT scaled scores. The data suggests that a combined approach, integrating ChatGPT with teacher instruction, might help EFL learners achieve a shift from CEFR levels B1 to C1.

Second, observational data indicate that the role of teachers remains indispensable and cannot be replaced by ChatGPT. Three areas stand out: understanding writing criteria, fostering critical thinking, and providing ethical guidance on academic writing. This suggests that, although ChatGPT can produce a written product, it cannot effectively scaffold the writing process—this is a task that only teachers can fulfill. ChatGPT is unable to generate or clarify opinions and reasoning related to

emotions without precise prompts, nor can it organize disjointed ideas. Thus, when ChatGPT fails to help learners comprehend writing criteria, it is also unable to assist them in resolving writing issues.

Finally, it is recommended by the author of this study that ChatGPT be integrated into EFL writing curricula, but under teacher supervision. To achieve effective implementation, ongoing practice and professional development in AI integration for writing instruction are essential. This will contribute to enhanced professionalism in the development of technology-driven education.

CONFLICTS OF INTEREST

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

FUNDING STATEMENT

No funding was received for this work.

PROCESS DATES

10, 2024

This manuscript was initially received for consideration for the journal on 09/22/2024, revisions were received for the manuscript following the double-anonymized peer review on 10/18/2024, the manuscript was formally accepted on 10/23/2024, and the manuscript was finalized for publication on 10/31/2024

CORRESPONDING AUTHOR

Correspondence should be addressed to Sirin Sawangwan; sirin_sa@rmutto.ac.th

REFERENCES

Aithal, P. S., & Aithal, S. (2023). Application of ChatGPT in higher education and research—A futuristic analysis. *International Journal of Applied Engineering and Management*, *7*(3), 168–194. DOI: 10.47992/IJAEML.2581.7000.0193

Alharbi, W. (2024). Mind the gap, please! Addressing the mismatch between teacher awareness and student AI adoption in language education within higher education. *International Journal of Computer-Assisted Language Learning and Teaching*, *14*(1), 1–28. DOI: 10.4018/IJCALLT.351245

Apridayani, A., Han, W., & Sakkanayok, K. (2024). Enhancing English writing competence in higher education: A comparative study of teacher-only assessment versus teacher and student self-assessment approaches. *Asian-Pacific Journal of Second and Foreign Language Education*, *9*(1), 37. DOI: 10.1186/s40862-024-00263-3

Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*, *7*(1), 52–62. DOI: 10.61969/jai.1337500

Baskara, R., & Mukarto, M. (2023). Exploring the implications of ChatGPT for language learning in higher education. *Indonesian Journal of English Language Teaching and Applied Linguistics*, *7*(2), 343–358. DOI: 10.21093/ijeltal.v7i2.1387

Berriche, L., & Larabi-Marie-Sainte, S. (2024). Unveiling ChatGPT text using writing style. *Heliyon*, *10*(12), 1–19. DOI: 10.1016/j.heliyon.2024.e32976 PMID: 38984302

Bigelow, M., & Watson, J. (Eds.). (2013). *The Routledge handbook of second language acquisition*. Routledge.

Bin-Hady, W. R. A., Al-Kadi, A., Hazaea, A., & Ali, J. K. M. (2023). Exploring the dimensions of ChatGPT in English language learning: A global perspective. *Library Hi Tech*, 28–39. DOI: 10.1108/LHT-05-2023-0200

ChatGPT. (2024, October 11). In *Wikipedia.* https://en.wikipedia.org/wiki/ChatGPT

Chen, X. L., Zou, D., Xie, H. R., & Su, F. (2021). Twenty-five years of computer-assisted language learning: A topic modeling analysis. *Language Learning & Technology, 25*(3), 151–185. DOI: 10125/73454

Chiu, T. F. K., Moorhouse, B. L., Chai, C. S., & Ismailov, M. (2023). Teacher support and student motivation to learn with artificial intelligence (AI) based chatbot. *Interactive Learning Environments*, *41*(1), 1–34. DOI: 10.1080/10494820.2023.2172044

Crystal, D. (2004). *Language and the internet*. Cambridge University Press.

Education First. (2023, February 23). *The world's largest rankings of countries and regions by English skills.* https://www.ef.com/wwen/epi/

Education First. (2024, April 30). *Converting from one English Level to Another.* https://doi.org. efset.org/english-score/#converting-from-one-english-level-to-another

Educational Testing Service. (2023, May, 27). *TOEFL Criteria Marking Scheme*. https://leap scholar.com/blog/toefl-writing-score-2/

Fitria, T. N. (2023). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. *Journal of English Language Teaching ELT FORUM*, *12*(1), 44–57. DOI: 10.15294/elt.v12i1.64069

Flower, L., & Hayes, J. R. (1980). Identifying the organisation of writing process. In Gregg, L. W., & Steinberg, E. R. (Eds.), *Cognitive processes in writing* (pp. 3–30). Erlbaum.

Hong, W. C. H. (2021). Macao Secondary School EFL teachers' perspectives on written corrective feedback: Rationales and constraints. *Journal of Educational Technology and Innovation*, *1*(4), 1–13. DOI: 10.61414/jeti.v1i1.17

Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation*, *5*(1), 37–45. DOI: 10.61414/jeti.v5i1.103

Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Journal of Educational Technology & Society*, *26*(1), 112–131. DOI: 10.30191/ETS.202301_26(1).0009

International Society for Technology in Education. (2024, October 11). *ISTE standards: For educators*. https://doi.org/iste.org/standards/educators

Javaid, M., Haleem, A., Singh, R. P., Khan, S., & Khan, I. H. (2023). Unlocking the opportunities through ChatGPT Tool towards ameliorating the education system. *Bench Council Transactions on Benchmarks. Standards and Evaluations*, *3*(2), 1–12. DOI: 10.1016/j.tbench.2023.100115

LeapScholar. (2024, September 28). *TOEFL writing score: Latest scoring system & marking scheme of 2023 explained*. https://doi.org.leapscholar.com/blog/toefl-writing-score-2/

Levy, M. (1997). *Computer-assisted language learning: Context and conceptualization*. Clarendon Press. DOI: 10.1093/oso/9780198236320.001.0001

Levy, M. (2009). Technologies in use for second language learning. *Modern Language Journal*, *93*(s1), 769–782. DOI: 10.1111/j.1540-4781.2009.00972.x

Li, X., Li, B., & Cho, S. J. (2023). Empowering Chinese language learners from low-income families to improve their Chinese Writing with ChatGPT's assistance afterschool. *Languages (Basel, Switzerland)*, *8*(4), 238–254. DOI: 10.3390/languages8040238

Lingard, L. (2023). Writing with ChatGPT: An illustration of its capacity, limitations & implications for academic writers. *Perspectives on Medical Education*, *12*(1), 261–270. DOI: 10.5334/pme.1072 PMID: 37397181

Masoudi, H. A. (2024). Effectiveness of ChatGPT in improving English writing proficiency among non-native English speakers. *International Journal of Educational Sciences and Arts*, *3*(4), 62–84. DOI: 10.59992/IJESA.2024.v3n4p2

Meniado, J. C., Huyen, D. T. T., Panyadilokpong, N., & Lertkomolwit, P. (2024). Using ChatGPT for second language writing: Experiences and perceptions of EFL learners in Thailand and Vietnam. *Computers and Education: Artificial Intelligence*, *7*, 1–9. DOI: 10.1016/j.caeai.2024.100313

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. Jossey-Bass.

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice*, *20*(2), 1–24. DOI: 10.53761/1.20.02.07

Reuters. (2023, January 27). *Top French university bans use of ChatGPT to prevent plagiarism*. https://doi.reuters.com/technology/top-french-university-bans-use-chatgpt-prevent-plag iarism-2023-01-27/

Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning and Teaching*, *6*(1), 342–363. DOI: 10.37074/jalt.2023.6.1.9

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*, 1260843. DOI: 10.3389/fpsyg.2023.1260843 PMID: 38162975

St-Onge, C., Ouellet, K., Lakhal, S., Dubé, T., & Marceau, M. (2022). COVID-19 as the tipping point for integrating e-assessment in higher education practices. *British Journal of Educational Technology*, *53*(2), 349–366. DOI: 10.1111/bjet.13169 PMID: 34898680

Swain, M. (1995). The output hypothesis: Theory and research. In Hilkel, E. (Ed.), *Handbook of research in second language teaching and learning* (pp. 495–508). Routledge.

Teaching English to Speakers of Other Languages International Association. (2024, April 5). *WAESOL online: ChatGPT: An assistant to enhance teaching & curriculum.* https://doi.tesol.org/calendar/event-listing/waesol-online-chatgpt-an-assistant-to-enhance-teaching-curriculum/

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huangand, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, *10*(1), 15. DOI: 10.1186/s40561-023-00237-x

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

Wang, Y., & Kruk, M. (2024). Modeling the interaction between teacher credibility, teacher confirmation, and English major students' academic engagement: A sequential mixed-methods approach. *Studies in Second Language Learning and Teaching*, *14*(2), 235–265. DOI: 10.14746/ssllt.38418

Warschauer, M. (1996). Computer-assisted language learning: An introduction. In Foto, S. (Ed.), *Multimedia language teaching* (pp. 3–20). Logos.

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies*, *28*(11), 13943–13967. DOI: 10.1007/s10639-023-11742-4