Arab World English Journal
INTERNATIONAL PEER REVIEWED JOURNAL ISSN: 2229-9327

Arab World English Journal (AWEJ) Volume 16. Number 2, June 2025 Pp.153-167
DOI: https://dx.doi.org/10.24093/awej/vol16no2.9

Impact of Flipped Classroom on Developing Critical Thinking Skills in Chinese EFL Undergraduates' Argumentative Essays: A Pilot Study

Huang Qi
School of Distance Education, Universiti Sains Malaysia, Penang, Malaysia
Email: huangqi@student.usm.my

Siti Soraya Lin Abdullah Kamal
School of Distance Education, Universiti Sains Malaysia, Penang, Malaysia
Corresponding Author**:** sorayalin@usm.my

Hicham Lahlou
School of Humanities, Universiti Sains Malaysia, Penang, Malaysia
Email: hicham@usm.my

Received:03/01/2025 Accepted:05/31/2025 Published:06/26/2025

Abstract
This research adopts a quantitative-qualitative mixed-methods approach to examine the role of flipped classroom pedagogy in fostering critical thinking skills in English argumentative essays among Chinese EFL undergraduates. Specifically, the pilot study compares the mean scores of the pretest and posttest, focusing on the performance in English argumentative essays between the experimental group (N=32) and the control group (N=28). It investigates the impact of the flipped classroom on developing critical thinking skills in argumentative essays. It explores the experiences and perceptions of Chinese EFL learners regarding how the flipped classroom enhances their critical thinking skills in such essays. The ultimate goal is to assist these students in improving their writing quality, achieving better academic outcomes, and enhancing their competitiveness in future job markets. The significance of this research lies in its potential to promote the adoption of the flipped classroom as an effective pedagogical tool for supporting Chinese EFL undergraduates. The findings indicate that the experimental group's mean scores steadily increased throughout the intervention, providing robust evidence of the effectiveness of the flipped classroom approach in fostering critical thinking skills in Chinese EFL undergraduates' argumentative essays writing. We hope this study will inspire more researchers to propose meaningful pedagogical implications and recommendations to improve teaching and learning processes within EFL contexts.

Keywords: Argumentative essay writing, critical thinking skills, Chinese EFL undergraduates, flipped classroom, mixed-method research
How to cite: Qi, H., Kamal, S. S.L. A., & Lahlou, H. (2025). Impact of Flipped Classroom on Developing Critical Thinking Skills in Chinese EF Undergraduates' Argumentative Essays: A Pilot Study. *Arab World English Journal, 16*(2):153-167.
https://dx.doi.org/10.24093/awej/vol16no2.9

Introduction

In an increasingly interconnected and competitive global landscape, proficiency in the English language has emerged as an essential skill, particularly for non-native speakers residing in countries such as China. For Chinese undergraduates, critically analysing information and articulating coherent arguments in English is crucial for achieving academic excellence and securing promising career opportunities within today's globally integrated workforce (Zhang & Kim, 2018). However, the traditional approach to teaching English as a Foreign Language (EFL) prevalent throughout China tends to emphasize rote memorization techniques alongside exam-oriented instruction (Ren & Wang, 2018). This methodology fails to foster the fundamental proficiency required for advanced scholarly pursuits or professional endeavours that require astute critical thinking and persuasive written communication skills.

As the demand for effective language instruction continues to rise, it becomes necessary to investigate innovative pedagogical methods. The flipped classroom (FC) model, defined by a reversed learning structure—transforming traditional approaches both upside down and inside out (Fulton, 2012)—presents a distinctive opportunity to enhance students' critical thinking skills (CTS). Specifically, teachers have transitioned conventional in-class activities to home-based tasks and transferred typical homework assignments to engaging classroom activities (Ma, 2023). Despite the growing global popularity of the FC model, its implementation in Chinese EFL contexts remains insufficiently examined. It is particularly pertinent given the ongoing academic discourse regarding the effectiveness of traditional versus innovative teaching methods in fostering CTS (Alsaleh, 2020). Moreover, there is a pressing need to address the challenges Chinese EFL students encounter in enhancing their critical thinking abilities and English proficiency in argumentative essays, both essential for academic institutions and employers.

Literature Review

"When Chinese students in the foreign language department encounter arguments and need to reason, they often experience cognitive inertia and struggle with articulation while writing articles, listening to lectures, and asking questions" (Huang, 1998, as cited in Wu, 2011, p.15). It is uncommon to find one or two commendable papers among a pool of twenty or thirty papers; most of these papers exhibit various issues, and frequently their titles reveal substantial deficiencies in critical thinking (Huang, 2010). Wu (2011) underscored that the absence of critical thinking directly engenders a multitude of drawbacks, thereby impeding the progress of a nation in the era of globalisation.

The curriculum in China emphasises improving language skills through imitation, memorisation, and repetitive drilling rather than developing a systematic understanding of the English language and promoting critical thinking and innovative abilities. There are defects concerning the content of writing teaching. What students receive from their teachers is general instructions about English language writing, or specific genre-based guidance, like narration, description, exposition, and argumentation. The education they receive in university primarily focuses on test-taking strategies for writing, which falls short of adequately preparing them to meet the demands of the job market or future competition in their professional careers.

A fundamental goal of education is to cultivate students' independent thinking ability to distinguish right from wrong, which is the essence of critical thinking. Argumentative essay writing is closely related to critical thinking and is regarded as the highest form of critical thinking because it requires expressing views, supporting arguments, and forming conclusions. The results of

Sharadgah et al.'s study underscore the significance of utilizing argumentative essays as an effective tool for both instructing and evaluating CTS (2019). Therefore, argumentative essay writing is also called critical writing. To get rid of the test-oriented English teaching and learning (Jin & Fan, 2011) as well as being away from writing an introduction of personal feelings or experience sharing due to the lack of critical thinking (Tian, 2021), Chinese EFL undergraduate students need comprehensive and sound instruction from teachers to improve argumentative essay writing capabilities.

Technology changes life and also changes people's ways of learning. Students born in 2000 and beyond are digital natives exposed to technology from an early age. Therefore, educators must keep up with technological advancements to effectively teach in this tech-savvy environment (Ahmad et al., 2019). Under these circumstances, the most urgent task is to use mobile technology properly in English instruction. Integrating the new mobile technology with a novel teaching design is a good solution. Moreover, as technology advances into the era of artificial intelligence, researchers and educators are increasingly inquiring into how it can facilitate or impede language learning, independent writing abilities and CTS (Avsheniuk et al., 2024; Pitychoutis, 2024).

As one of the learning innovations, people regard FC as a preeminent learning model (Namaziandost & Çakmak, 2020). It enables students to preview the lesson at home beforehand and do homework in class the day after (Fulton, 2012). Many Chinese university and college faculty members advocated FC in their online instruction to replace the traditional teaching approach. They like to adopt it specifically due to its innovative operation and accessibility compared with other methods. In comparison to conventional teaching methods, the FC approach has been widely appreciated by students (Ramnanan & Pound, 2017; Ramirez et al., 2022; Balasubramanian et al., 2022), as it is expected to enhance and foster students' intrinsic motivation (Howell, 2021) while also being believed to have the potential for promoting students' CTS development (Nugraheni et al., 2022). However, it remains in an underdeveloped state (Wang, Huang, Schunn, Zou & Ai, 2019).

Current English language teaching in China prioritizes language mechanics over CTS, leading to EFL undergraduates' academic writing deficiencies. While much research emphasizes the importance of CTS, few studies investigate how to specifically develop these skills within the context of English argumentative essay writing for Chinese EFL undergraduates. As a result of the inadequacy of traditional teaching methods in Chinese EFL education, as mentioned above and the underutilisation of FC in Chinese EFL classrooms, this research aims to address the gap by examining the effectiveness of implementing the FC approach in fostering CTS in Chinese EFL undergraduates' argumentative essays.

Method

The present study employed a quasi-experimental design (Bond, 2020; Junianti et al., 2020; Atwa et al., 2022) in its quantitative investigation, followed by conducting interviews in the qualitative research phase, thus adopting a mixed-methods approach (Liao & Li, 2017; Ahmed, 2019). This study lies in "QUAN→qual Capital letters indicate theoretical drive or priority (core methods) given in a study; lower case indicates (supplemental methods)" (Tashakkori & Teddlie, 2010, p. 57) .

The research team designed the study to evaluate the effectiveness of FC in developing CTS of Chinese EFL students in argumentative essay writing at the undergraduate level. The subjects of

this study were undergraduates in the same grade to fulfil the purpose of homogeneity (Adhami & Taghizadeh, 2022). The research team divided them into two groups, experimental and control group (CG) (Sharadgah et al., 2019; Nejmaoui, 2019; Song & Wang, 2022; Hassan et al., 2023). Although instructed using distinct methodologies, both groups completed identical assessments. The research team systematically designed and administered a pretest and posttest (Atwa et al., 2022) and structured interviews (Abrami et al., 2015).

Both groups participated in the pretest and posttest, while only the experimental group (EG) received the intervention (Creswell, 2003). The research team conducted the pretest and posttest to collect data from Chinese EFL undergraduates' practice essays and assess the students' progress before and after the research. The research was going to study FC and CTS within the paradigm. The purpose of the pretest was to assess students' original CTS in argumentative essay writing, and the posttest was to determine undergraduates' CTS after the treatment teaching process. Each test would involve two groups completing an identical writing task. In total, the pilot study comprised three tests.

The writing tests had been administered and assessed, with each argumentative essay receiving a score ranging from one to thirteen marks based on its academic writing quality in both groups. Student participants should write the essays in paragraphs. Its requirements include a title, an introductory paragraph, two to three body paragraphs, and a concluding paragraph. The research aims to evaluate student participants' writing speed and language proficiency. Therefore, the word count should fall within the range of 150 to 350 (Xie, 2023) while adhering to a time limitation of forty minutes. Then, two qualified and experienced teachers will grade these written essays to ensure validity. Plagiarism was strictly prohibited.

After the posttest, the team conducted structured individual interviews to gather qualitative data. In the pilot study, the researcher interviewed six students from two groups representing different language competencies separately to learn their views about the treatment and traditional teaching method. This approach allowed the researcher to comprehensively comprehend the experiences and perspectives regarding FC among student participants with varying language proficiency and CTS.

Instruments

The VQGOIP Teaching Guideline and Lesson Plan were utilized in the instructional process to enhance undergraduate students' CTS by implementing FC within the context of argumentative essays—a panel of three experts recruited by the research team before the study had validated the materials. The research objectives were pursued by collecting quantitative and qualitative data, enabling a comprehensive analysis of multiple facets. The research team used the IBM Statistical Package for the Social Sciences (SPSS) software version 29 to obtain quantitative data. The team primarily obtained qualitative data through a structured individual interview with student participants from both groups.

The delivery and assessment within an FC necessitate addressing three key aspects: "the educational platform, content organization, and planned assessments" (Zappe & Litzinger, 2017, as cited in Petre, 2020, p.8). The research utilized "Xuexitong" as the learning platform, employed the FC Teaching Guideline for content organization in teaching, and incorporated pretest and

posttest assessments to evaluate the effectiveness of FC in enhancing CTS of English argumentative essay writing among Chinese EFL undergraduates. Apart from that, the research team compared the mean scores of the EG with those of the CG to assess any disparities in their learning outcomes.

The researcher developed the VQGOIP teaching guideline, drawing on extensive references. Figure 1 illustrates the VQGOIP teaching guideline. To ensure the credibility of this teaching guideline, which the researcher designed, validation was obtained from an English lecturer with over a decade of experience in higher education prior to the pilot study, following which the research phase would commence. The researcher designed this teaching guideline for the FC approach, and its associated lesson plans were tailored explicitly for two distinct groups. Only the EG was involved in the FC teaching intervention, whereas the CG used traditional teaching methods.

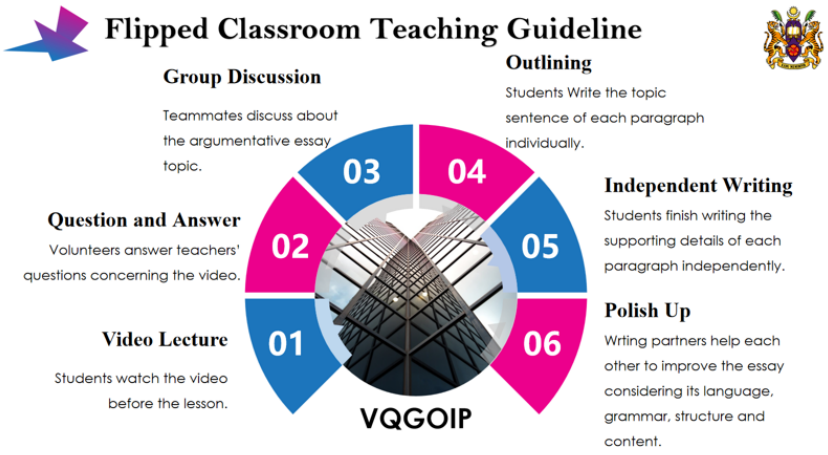


Figure 1. The FC VQGOIP Teaching Guideline

CG utilized lesson plans specifically designed for the traditional lecturing method, using the textbook edited by Zou (2019). Both groups employed distinct lesson plans for instruction. Table 1 illustrates the differential interventions in the two groups.

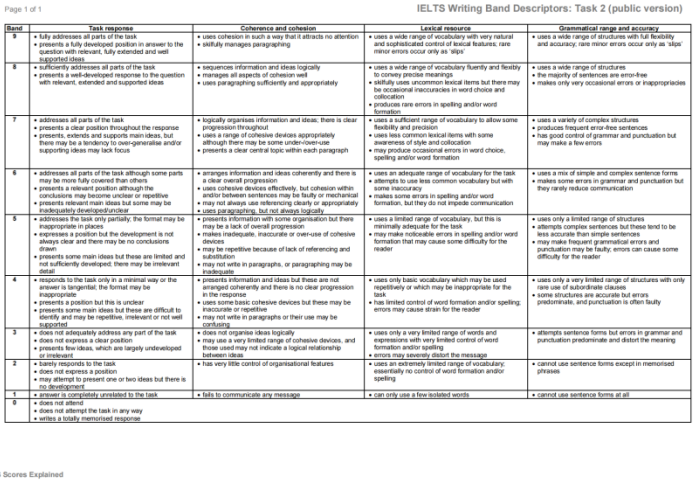
Table 1. *Differential Interventions in the Two Groups*

Group	Teaching Methods	Lesson Plan
EG	FC Utilizing the VQGOIP Teaching Guideline	LP for EG
CG	Traditional Lecturing Method and a textbook	LP for CG

Rating Rubrics for the Pilot Study

The assessment criteria for the argumentative essays had integrated elements from the *IELTS Writing Band Descriptors: Task 2 (public version)* and *The Holistic Critical Thinking Scoring Rubric* (HCTSR) developed by Facione and Facione (1994). Given that writing skills encompass two dimensions—students' English language proficiency and their CTS—the rating

process consisted of two sequential stages. Initially, raters evaluated students' argumentative essays based on the IELTS Writing Band Descriptors to assess their language proficiency. Subsequently, they evaluated the essays using the *HCTSR* (Facione & Facione, 1994) to analyze CTS. Consequently, students' total scores range from one to thirteen. The maximum score for academic argumentative writing is nine, reflecting their English language proficiency, with an additional four points awarded based on their CTS level. Within this process, the CTS grade served as an additive component. The minimum score for English proficiency in writing is zero, with additional points granted according to the CTS level. Raters were required to sum the scores from both evaluation stages. *The IELTS Writing Band Descriptors: Task 2 (public version)* are illustrated in Figure 2, while Figure 3 presents the *HCTSR*.



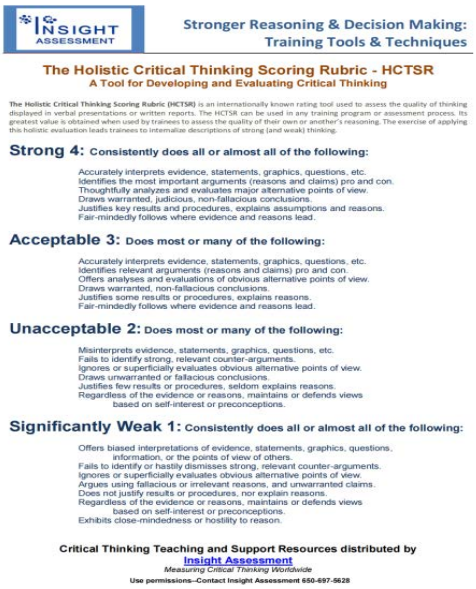
Page 1 of 1

IELTS Writing Band Descriptors: Task 2 (public version)

Band	Task response	Coherence and cohesion	Lexical resource	Grammatical range and accuracy
9	• fully addresses all parts of the task • presents a fully developed position in answer to the question with relevant, fully extended and well supported ideas	• uses cohesion in such a way that it attracts no attention • skilfully manages paragraphing	• uses a wide range of vocabulary with very natural and sophisticated control of lexical features; rare minor errors occur only as 'slips'	• uses a wide range of structures with full flexibility and accuracy; rare minor errors occur only as 'slips'
8	• sufficiently addresses all parts of the task • presents a well-developed response to the question with relevant, extended and supported ideas	• sequences information and ideas logically • manages all aspects of cohesion well • uses paragraphing sufficiently and appropriately	• uses a wide range of vocabulary fluently and flexibly to convey precise meanings • skilfully uses uncommon lexical items but there may be occasional inaccuracies in word choice and collocation • produces rare errors in spelling and/or word formation	• uses a wide range of structures • the majority of sentences are error-free • makes only very occasional errors or inappropriacies
7	• addresses all parts of the task • presents a clear position throughout the response • presents, extends and supports main ideas, but there may be a tendency to over-generalise and/or supporting ideas may lack focus	• logically organises information and ideas; there is clear progression throughout • uses a range of cohesive devices appropriately although there may be some under-/over-use • presents a clear central topic within each paragraph	• uses a sufficient range of vocabulary to allow some flexibility and precision • uses less common lexical items with some awareness of style and collocation • may produce occasional errors in word choice, spelling and/or word formation	• uses a variety of complex structures • produces frequent error-free sentences • has good control of grammar and punctuation but may make a few errors
6	• addresses all parts of the task although some parts may be more fully covered than others • presents a relevant position although the conclusions may become unclear or repetitive • presents relevant main ideas but some may be inadequately developed/unclear	• arranges information and ideas coherently and there is a clear overall progression • uses cohesive devices effectively, but cohesion within and/or between sentences may be faulty or mechanical • may not always use referencing clearly or appropriately • uses paragraphing, but not always logically	• uses an adequate range of vocabulary for the task • attempts to use less common vocabulary but with some inaccuracy • makes some errors in spelling and/or word formation, but they do not impede communication	• uses a mix of simple and complex sentence forms • makes some errors in grammar and punctuation but they rarely reduce communication
5	• addresses the task only partially; the format may be inappropriate in places • expresses a position but the development is not always clear and there may be no conclusions drawn • presents some main ideas but these are limited and not sufficiently developed; there may be irrelevant detail	• presents information with some organisation but there may be a lack of overall progression • makes inadequate, inaccurate or over-use of cohesive devices • may be repetitive because of lack of referencing and substitution • may not write in paragraphs, or paragraphing may be inadequate	• uses a limited range of vocabulary, but this is minimally adequate for the task • may make noticeable errors in spelling and/or word formation that may cause some difficulty for the reader	• uses only a limited range of structures • attempts complex sentences but these tend to be less accurate than simple sentences • may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader
4	• responds to the task only in a minimal way or the answer is tangential; the format may be inappropriate • presents a position but this is unclear • presents some main ideas but these are difficult to identify and may be repetitive, irrelevant or not well supported	• presents information and ideas but these are not arranged coherently and there is no clear progression in the response • uses some basic cohesive devices but these may be inaccurate or repetitive • may not write in paragraphs or their use may be confusing	• uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task • has limited control of word formation and/or spelling; errors may cause strain for the reader	• uses only a very limited range of structures with only rare use of subordinate clauses • some structures are accurate but errors predominate, and punctuation is often faulty
3	• does not adequately address any part of the task • does not express a clear position • presents few ideas, which are largely undeveloped or irrelevant	• does not organise ideas logically • may use a very limited range of cohesive devices, and those used may not indicate a logical relationship between ideas	• uses only a very limited range of words and expressions with very limited control of word formation and/or spelling • errors may severely distort the message	• attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning
2	• barely responds to the task • does not express a position • may attempt to present one or two ideas but there is no development	• has very little control of organisational features	• uses an extremely limited range of vocabulary; essentially no control of word formation and/or spelling	• cannot use sentence forms except in memorised phrases
1	• answer is completely unrelated to the task	• fails to communicate any message	• can only use a few isolated words	• cannot use sentence forms at all
0	• does not attend • does not attempt the task in any way • writes a totally memorised response			

IELTS Scores Explained

Figure 2. IELTS Writing Band Descriptors: Task 2 (public version)



INSIGHT ASSESSMENT

Stronger Reasoning & Decision Making:
Training Tools & Techniques

The Holistic Critical Thinking Scoring Rubric - HCTSR
A Tool for Developing and Evaluating Critical Thinking

The Holistic Critical Thinking Scoring Rubric (HCTSR) is an internationally known rating tool used to assess the quality of thinking displayed in verbal presentations or written reports. The HCTSR can be used in any training program or assessment process. Its greatest value is obtained when used by trainees to assess the quality of their own or another's reasoning. The exercise of applying this holistic evaluation leads trainees to internalize descriptions of strong (and weak) thinking.

Strong 4: Consistently does all or almost all of the following:

Accurately interprets evidence, statements, graphics, questions, etc.
Identifies the most important arguments (reasons and claims) pro and con.
Thoughtfully analyzes and evaluates major alternative points of view.
Draws warranted, judicious, non-fallacious conclusions.
Justifies key results and procedures, explains assumptions and reasons.
Fair-mindedly follows where evidence and reasons lead.

Acceptable 3: Does most or many of the following:

Accurately interprets evidence, statements, graphics, questions, etc.
Identifies relevant arguments (reasons and claims) pro and con.
Offers analyses and evaluations of obvious alternative points of view.
Draws warranted, non-fallacious conclusions.
Justifies some results or procedures, explains reasons.
Fair-mindedly follows where evidence and reasons lead.

Unacceptable 2: Does most or many of the following:

Misinterprets evidence, statements, graphics, questions, etc.
Fails to identify strong, relevant counter-arguments.
Ignores or superficially evaluates obvious alternative points of view.
Draws unwarranted or fallacious conclusions.
Justifies few results or procedures, seldom explains reasons.
Regardless of the evidence or reasons, maintains or defends views based on self-interest or preconceptions.

Significantly Weak 1: Consistently does all or almost all of the following:

Offers biased interpretations of evidence, statements, graphics, questions, information, or the points of view of others.
Fails to identify or hastily dismisses strong, relevant counter-arguments.
Ignores or superficially evaluates obvious alternative points of view.
Argues using fallacious or irrelevant reasons, and unwarranted claims.
Does not justify results or procedures, nor explain reasons.
Regardless of the evidence or reasons, maintains or defends views based on self-interest or preconceptions.
Exhibits close-mindedness or hostility to reason.

Critical Thinking Teaching and Support Resources distributed by
Insight Assessment
Measuring Critical Thinking Worldwide
Use permissions--Contact Insight Assessment 650-697-5628

Figure 3. The Holistic Critical Thinking Scoring Rubric - HCTSR

Participants

Sample Selection

Sampling Technique for the Quantitative Study

We have previously noted the challenges associated with academic English writing courses and critical thinking in English. Consequently, the research team selected English majors as the sample group. The team employed a purposive sampling technique among English majors at the institution. Only student participants who demonstrate fluency in English are eligible for inclusion. Furthermore, the class with a higher mean pretest score was designated as the EG, whereas the class with a lower mean pretest score was the CG.

Sampling Technique for the Qualitative Study

The selection of student participants for the interviews reflects the representative nature of the sample, thereby ensuring a typical case in qualitative studies. Participants were chosen based on their posttest scores, with two students selected from the advanced level, two from the intermediate level, and two from the basic level within the designated group. The research team conducted the interviews, during which participants needed to respond to a series of specifically designed questions in sequential order. Appendix A has outlined the interview questions.

The student participants comprised thirty-two and twenty-eight members (Fatemipour & Jafari, 2015), respectively, from each cohort. The teacher participants had two individuals assigned as raters while the remaining two served as instructors in their respective groups. None had previously resided in countries where English is the primary language (Alpat, 2019). The participants had enrolled in a one-semester English writing course, with their ages ranging from 18 to 21 (Sharadgah et al., 2019).

Procedures

Data Collection Procedures

Collecting quantitative data involved conducting a pretest, evaluating one additional argumentative essay, and administering a posttest. As previously mentioned, the class with a higher mean score in the pretest within the two formed the EG (Female=27; Male=5). Conversely, students of the other class with a lower mean score in the pretest, also participating in academic English writing courses, constituted the CG (Female=24; Male=4). In this study, Teacher A provided intervention instruction to the EG, whereas Teacher B conducted instruction for the CG. The procedure was maintained consistently for both groups, with the sole difference lying in their instructional methods. Despite the distinct methodologies employed for instruction, both groups had experienced identical assessments.

The face-to-face structured individual interview involved the researcher conducting interviews with the student participants. The session included a minimum of six respondents and one interviewer. These respondents were student participants responsible for responding to the upcoming inquiries, while the interviewer was the researcher who assumed the role of posing questions. The inquiries were open-ended, and the interviewer made utmost efforts to elicit the interviewees' genuine thoughts. The two parties strove to complete the task through interactive communication. Data saturation would determine the sample size of the individual interview. The data collection was discontinued upon reaching the point of data saturation, as no additional information could be obtained once it had been achieved. In the interview, the team instructed participants to capture photographs or record audio (i.e., engage in photo elicitation) (Creswell,

2003). The researcher would utilize these photographs or recordings for subsequent transcription and analysis.

The Quantitative Study

Results of the Quantitative Study

The pilot study lasted three weeks, during which the research team organized a weekly test to assess student participants' progress. Both groups conducted identical tests throughout the study. Table 2 and Figure 4 present the mean scores of both groups during the pilot study. In Figure 4, the blue line represents the EG, while the red line indicates the CG. In summary, the mean scores for both groups fell within the range of 4.9 to 6.9. According to the graphs, the mean scores consistently exhibit an upward trend throughout the study, particularly for EG. In contrast, CG's mean scores initially show an increase followed by a subsequent decrease. Despite greater fluctuations observed in the latter group, EG consistently outperforms CG.

Table 2. *Mean Score of the Pilot Study*

Group	Mean Score of the Pretest	Mean Score of Argumentative Essay Writing Practice	Mean Score of the Posttest
EG	5.59	6.606	6.844
CG	4.95	6.586	6.071

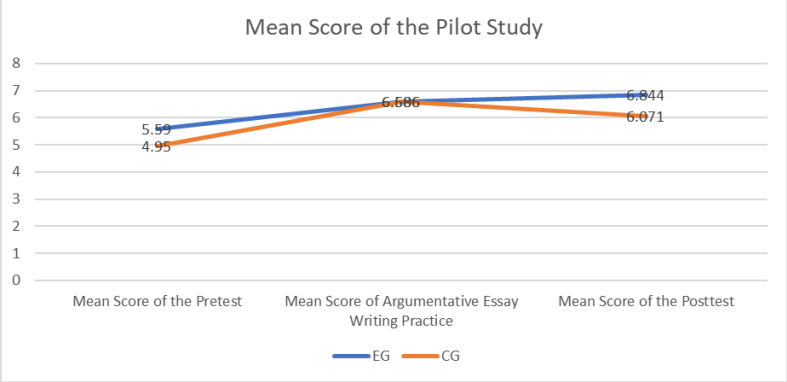


Figure 4. Line Graph Depicting Mean Scores from the Pilot Study

Discussion for the Quantitative Study

The research team employed a paired-samples t-test to assess the efficacy of the intervention on students' scores for English argumentative essays. The results, as detailed in Tables 3 and 4, revealed a statistically significant difference in the EG's scores before ($M = 5.59$, $SD = 1.66$) and after the intervention ($M = 6.84$, $SD = 1.04$), with $t(31) = -4.902$, $p < 0.05$ (Hinton et al., 2014). It indicates that the intervention was associated with a substantial improvement in scores. According to Tables 3 and 5, during the pilot study, the mean score of the EG increased by 1.25 points (from

5.594 to 6.844), while the mean score of the CG increased by 1.125 points (from 4.946 to 6.071). Moreover, the mean scores for the EG were consistently higher than those of the CG in both the pretest (M = 4.95, SD = 1.72) and posttest (M = 6.07, SD = 1.50). Given that no intervention was administered to the CG, this comparison provides further evidence supporting the effectiveness of the intervention in the EG.

Table 3. *Output of EG*

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	pretest	5.594	32	1.6580	.2931
	posttest	6.844	32	1.0429	.1844

Table 4. *Output of the Paired Samples T-Test*

Paired Samples Test										
		Paired Differences					t	df	Significance	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				One-Sided p	Two-Sided p
					Lower	Upper				
Pair 1	pretest - posttest	-1.2500	1.4424	.2550	-1.7701	-.7299	-4.902	31	<.001	<.001

Table 5. *Output of CG*

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	pretest	4.946	28	1.7178	.3246
	posttest	6.071	28	1.5013	.2837

Results and Discussion of the Qualitative Study

Through structured face-to-face individual interviews, we gained insights into the perceptions and experiences of student participants regarding the cultivation of CTS via English argumentative essays in an FC setting. We will analyze the data from the following three perspectives:

Misconceptions Regarding FC and CTS

The misunderstanding of FC is significant. Four of the six student participants expressed erroneous concepts, while the remaining two were unfamiliar with the precise definition. Misconceptions often stem from prior experiences or perceptions, which implies that instructors may misinform individuals during their previous learning experiences. Some teachers have misunderstood the concept of FC, leading to students' misguidance. It is advantageous for educators to transition from traditional instructional methods to innovative approaches; however, it is imperative to implement these changes appropriately to prevent misleading students.

During the interview with N1, he mentioned, "My friends believe that critical thinking can sometimes be perceived as overly assertive or unwelcoming, leading to feelings of being targeted." This perception reflects a stereotype that may contribute to the current limited acceptance of CTS among some Chinese people. They believed that the individual utilizing CTS was targeting them. CTS is merely discussing and analyzing the issue reasonably; it does not pertain to any individual involved. We should evaluate numerous matters on their merits from multiple perspectives, independent of personal factors. Prejudice misleads. Therefore, the primary objective of our future main study is to emphasize the objectivity and scientific rigour of CTS.

Attitudes Toward Challenging Learning Contents

The learning attitudes of student participants correlate with their English proficiency levels and interest in learning. Specifically, students with higher proficiency levels, such as N1 and N4, tend to exhibit more positive attitudes. Additionally, students with a strong interest in learning English, like N2, demonstrate a more active approach to their studies. Although most students prefer active learning, those with lower proficiency levels, such as N3, may become overly reliant on their partners in pair work or teammates in teamwork. This dependence arises from the need for classmates' support, encouragement, and even teachers' stimulation. It aligns with Vygotsky's theory of the Zone of Proximal Development. The ZPD refers to the gap between students' ability to perform tasks with adult guidance or collaboration with peers and their capacity to solve problems independently (Ganapathy, 2011). Cooperative learning and collaborative writing (Doyin & Femiti, 2025) significantly benefit these students.

Nevertheless, there are also counterexamples, such as N6. Despite her struggling English proficiency in the posttest, her attitude remains exemplary. She dedicates significant time to practising specific language skills to improve, which exemplifies her persistence and tenacity. It highlights a strong will's crucial role in overcoming academic challenges. Active learning can address students' deficiencies in English proficiency.

Habits in Utilizing Technology

During the interview, N1 noted that his roommates were unfamiliar with the operating procedures for viewing micro-videos. Specifically, they did not know how to pause the videos when necessary, making it difficult to understand the content thoroughly. Although most interviewees were familiar with the operating procedures, N1's sharing highlighted the necessity for training student participants in the main study to enhance their comprehension of the videos. Moreover, student participants faced challenges adapting to the FC approach, as they had recently transitioned from high school and had not become accustomed to using smartphones and tablets for learning purposes. Previously, students used these devices for entertainment rather than academic study. Therefore, a three-week period is insufficient for participants to acclimate fully. The research team consequently decided to extend the duration of the main study to three months.

Conclusion

Based on the foregoing discussion, the performance of English argumentative essays in the EG is superior to that of the CG. This superiority was due to the EG's participation in a teaching intervention designed to enhance CTS through the FC approach during the academic English writing course. Therefore, implementing FC pedagogy has proven effective in developing CTS among Chinese EFL undergraduates, as evidenced by their improved performance in English argumentative essays and students' experiences within the interviews. Future research could explore the integration of CTS with FC across various genres of English writing.

Funding: This research is not funded.

Conflicts of Interest: The authors declare no conflicts of interest.

Authenticity: This manuscript is an original work.

Artificial Intelligence Statement: AI and AI-assisted technologies were not used.

About the authors

Huang Qi is currently a Ph.D. candidate at the School of Distance Education, Universiti Sains Malaysia, and concurrently serves as a lecturer in the English Department of the Foreign Language School at Shaoguan University, Guangdong Province, China. Her research focuses primarily on TESOL. She has over 16 years of teaching experience in English in Chinese higher education. ORCID ID : https://orcid.org/0009-0004-7271-1431

Siti Soraya Lin Abdullah Kamal obtained her doctorate in education from the University of Sheffield, UK. Her dissertation explored the social, cultural, and contextual factors influencing struggling readers of English as a second language among young learners in Malaysian classrooms. Her research interests include TESOL, language learning (early childhood and

primary education), e-learning, educational technology, distance learning, and qualitative studies. ORCID ID：https://orcid.org/0000-0002-2848-7039

Hicham Lahlou (PhD) is a senior lecturer at the English Language Studies Section of the School of Humanities, Universiti Sains Malaysia. He holds a doctorate in applied linguistics and has over two decades of experience teaching English and linguistics. His research interests include English linguistics, semantics, morphology, cognitive linguistics, and education.
ORCID ID：https://orcid.org/0000-0003-2796-9877

References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of educational research*, *85*(2), 275-314.

Adhami, N & Taghizadeh, M. (2022). Integrating inquiry-based learning and computer supported collaborative learning into flipped classroom: effects on academic writing performance and perceptions of students of railway engineering. *Computer Assisted Language Learning*, 1-31. https://doi.org/10.1080/09588221.2022.2046107

Ahmad, S. A., Yoke, S. K., Yunos, R. M., & Amin, J. M. (2019). Exploring lecturers' readiness for 21st century education in Malaysian higher learning institutions. *European Journal of Teaching and Education, 1*(1), 15-29. https://doi.org/10.33422/EJTE.2019.10.27

Ahmed, P. H. (2019). Major writing challenges experienced by EFL learners in Soran University. *Journal of University of Human Development, 5*(3), 120-126.

Alpat, M. F. (2019). *The effect of flipped learning-supported critical thinking instruction on the critical disposition and l2 writing skills* (Master's thesis, İstanbul Sabahattin Zaim Üniversitesi, Sosyal Bilimler Enstitüsü, Yabancı Diller Anabilim Dalı).

Alsaleh, N. J. (2020). Teaching Critical Thinking Skills: Literature Review. *Turkish Online Journal of Educational Technology-TOJET, 19*(1), 21-39.

Atwa, Z., Sulayeh, Y., Abdelhadi, A., Jazar, H. A., & Eriqat, S. (2022). Flipped Classroom Effects on Grade 9 Students' Critical Thinking Skills, Psychological Stress, and Academic Achievement. *International Journal of Instruction, 15*(2), 737-750.

Avsheniuk, N., Lutsenko, O., Svyrydiuk, T., & Seminikhyna, N. (2024). Empowering language learners' critical thinking: evaluating ChatGPT's role in English course implementation. *Arab World English Journal (AWEJ) Special Issue on ChatGPT*.

Balasubramanian, K., Sethuraman, V. K., & Ramachandran, B. (2022). A comparative study of students' perception of flipped classroom teaching and traditional classroom teaching among final-year medical students. *Journal of Current Research in Scientific Medicine, 8*(1), 52-58.

Bond, M. (2020). Facilitating student engagement through the flipped learning approach in K-12: A systematic review. *Computers & Education, 151,* 103819.

Creswell, J. W. (2003). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Second Edition.* SAGE Publications.

Doyin, Y., & Femiti, O. (2025). English Education in EFL Contexts: An Integrated Approach to Collaborative Writing at the University Level.

Facione, P. A., & Facione, N. C. (1994). The Holistic Critical Thinking Scoring Rubric-HCTSR. *A Tool for Developing and Evaluating Critical Thinking. URL: https://www.insightassessment.com/Resources/Teaching-Training-and-Learning-Tools/Holistic-Critical Thinking-Scoring-Rubric-HCTSR.*

Fatemipour, H., & Jafari, F. (2015). The effect of dynamic-assessment on the development of passive vocabulary of intermediate EFL learners. *Journal of Educational and Management Studies, 5*(1), 41-51.

Fulton, K. (2012). Upside down and inside out: Flip your classroom to improve student learning. *Learning & Leading with Technology, 39*(8), 12-17.

Ganapathy, M. (2011). The effects of the multiliteracies approach on the writing performance of EFL students in a Chinese school in Penang: Teacher and student perceptions [PhD thesis]. *Universiti Sains Malaysia.*

Hassan, A., Alawawda, M., Alzahrani, F., and Naz, N. (2023). Developing an ESP-Based Language Learning Environment to Help Students Improve Critical Thinking Skills in Written Output. I*nformation Sciences Letters, 12*(4), 1131-1140. https://digitalcommons.aaru.edu.jo/isl/vol12/iss4/31

Huang, Y. (1998). On the critical thinking deficiency. *Foreign Languages and their teaching, 7*(1), 1-19.

Huang, Y. (2010). The English Major Curriculum Must Be Overhauled Completely—On the "critical thinking deficiency" again. *Foreign Language World,* (1), 11-16.

Hinton, P. R., McMurray, I., & Brownlow, C. (2014). *SPSS explained.* Routledge.

Howell, R. A. (2021). Engaging students in education for sustainable development: The benefits of active learning, reflective practices and flipped classroom pedagogies. *Journal of Cleaner Production, 325,* 129318.

IELTS Writing Band Descriptors: Task 2 (public version) *https://ieltsimmigration.ir/wp-content/uploads/2020/03/Band-Descriptors-Task-2.pdf*

Jin, Y., & Fan, J. (2011). Test for English majors (TEM) in China. *Language Testing, 28*(4), 589-596.

Junianti, R., Pratolo, B. W., & Wulandari, A. T. (2020). The strategies of learning writing used by EFL learners at a higher education institution. Ethical Lingua: *Journal of Language Teaching and Literature, 7*(1), 64-73.

Liao, H., & Li, L. (2017). A mixed methods evaluation of college English writing: a case study in China. *The Asia-Pacific Education Researcher, 26*(6), 383-396.

Ma, Y. (2023). Exploration of flipped classroom approach to enhance critical thinking skills. *Heliyon, 9*(11).

Namaziandost, E., & Çakmak, F. (2020). An account of EFL learners' self-efficacy and gender in the Flipped Classroom Model. *Education and Information Technologies, 25*(5), 4041-4055.

Nejmaoui, N. (2019). Improving EFL Learners' Critical Thinking Skills in Argumentative Writing. *English Language Teaching*, *12*(1): 98-109. Canadian Centre of Science and Education.

Nugraheni, B. I., Surjono, H. D., & Aji, G. P. (2022). How can flipped classroom develop critical thinking skills? A literature review. *International Journal of Information and Education Technology, 12*(1), 82-90.

Petre, G. (2020). Using flipped classroom to facilitate cooperative learning implementation: An action research case study with photovoice. *Journal of Education Studies, 2*(2), 5-35.

Pitychoutis, K. M. (2024). Harnessing AI Chatbots for EFL Essay Writing: A Paradigm Shift in

Language Pedagogy. *Arab World English Journal (AWEJ) Special Issue on ChatGPT, April 2024: 197–209.* DOI: https://dx.doi.org/10.24093/awej/ChatGPT.13

Ramirez, S., Teten, S., Mamo, M., Speth, C., Kettler, T., & Sindelar, M. (2022). Student Perceptions and Performance in a Traditional, Flipped classroom, and Online Introductory Soil Science Course. *Journal of Geoscience Education, 70*(1), 130-141.

Ramnanan, C. J., & Pound, L. D. (2017). Advances in medical education and practice: student perceptions of the flipped classroom. *Advances in medical education and practice,* 63-73.

Ren, J., & Wang, N. (2018). Production-Oriented Approach and Its Implications for the Cultivation of Critical Thinking Skills in College English Instruction in Mainland China. *English language teaching, 11*(5), 33-38.

Sharadgah, T. A., Sa'di, R. A., & Ahmad, H. H. (2019). Promoting and Assessing EFL College Students' Critical Thinking Skills through Argumentative Essay Writing. *Arab World English Journal, 10 (4) 133-150.*

Song, Y. & Wang, S. (2022). The Application of Pragmatic Presupposition in Developing Critical Thinking of Writing. *Chinese Journal of Applied Linguistics, 45*(4), 583-595.

Tashakkori, A. & Teddlie, C. (2010). *SAGE Handbook of Mixed Methods in Social & Behavioural Research.* Second Edition. SAGE Publications.

Tian, H. (2021). *Critical Thinking and Writing.* Peking University Press.

Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes.* Cambridge: Harvard University Press.

Wang, Y., Huang, X., Schunn, C. D., Zou, Y., & Ai, W. (2019). Redesigning flipped classrooms: a learning model and its effects on student perceptions. *Higher Education, 78, 711-728.*

Wu, K. (2011). On Absence of Critical Thinking and Rhetorical Invention. *Chinese University Teaching, 10,* 15-17.

Xie, Q. (2023). Assessing source use: Summary vs. reading-to-write argumentative essay. *Assessing Writing, 57,* 100755.

Zappe, S. E., & Litzinger, T. A. (2016). Considerations when evaluating the classroom flip instructional technique. In *The flipped college classroom: Conceptualised and re-conceptualised* (pp. 51-62). Cham: Springer International Publishing.

Zhang, L., & Kim, S. (2018). Critical Thinking Cultivation in Chinese College English Classes. *English Language Teaching, 11*(8), 1-6.

Zou, S. (2019). *Successful Writing Students' Book 1: The Second Edition.* Shanghai Foreign Language Education Press.

Appendix A: Interview Questions

Number	Question
1	Do you know what Critical Thinking Skills are? What is the Flipped Classroom? Have you ever experienced these two before the instruction of the academic English writing course?
2	Do you like writing an English argumentative essay before attending the course? Then, after studying the course, have your opinions changed yet? Why or why not?

3	Do you agree with the saying that the Flipped Classroom has enhanced your Critical Thinking Skills concerning argumentative essay writing? Why said so? Or why not? Can you talk me through how the teacher went about using that?
4	Are you interested in the teacher's inquiries regarding video comprehension? Why or why not?
5	Which micro-video lesson is your favourite? Why?
6	Have you accomplished Critical Thinking Skills after the course? Which skill impresses you most: interpretation, analysis, evaluation, inference, explanation and self-regulation? Which one is the most difficult to utilise in your writing?
7	Which one do you prefer, individual work, pair work, or teamwork? How often do you usually cooperate with your classmates during the semester? What is your opinion about cooperative learning?
8	Do you think that Critical Thinking Skills are important? How about decision-making? Is it necessary for us to learn CTS at university? Why or why not?
9	How do you feel about group discussion in the class? Is it helpful to improve your Critical Thinking Skills? Why or why not?
10	What do you think of your teacher's instruction? Is it easy to understand? If the teaching content is difficult, how will you make it learnable? Active learning or passive learning, which way do you prefer?
11	How do you develop your Critical Thinking Skills in daily life? Can you give me some examples?
12	As regards the interview today, how do you perceive it? What are your thoughts on the questions asked during the interview? Is it easy, so-so, or difficult to answer these questions?
13	In general, would you have preferred an increased/decreased amount of lecture time from the teacher?
14	Can you share some positive/negative experiences of the flipped classroom to learn CTS? Why do you call these experiences positive or negative?