RESEARCH ARTICLE VOLUME 20, ISSUE 1, 93-110

The New English Teacher
ISSN 2985-0959 (Online)

# The Impact of Chaoxing-based Flipped Classroom on Critical Thinking Skills of English Majors in China: A Mixed-Methods Approach

***Shimeng Pan***
Graduate School of Business and Advanced Technology Management
Assumption University
88 Moo 8 Bang Na-Trad Km. 26, Bangsaothong, Samutprakarn, Thailand, 10570
Tel: +8613-3900-82211
Email: poemdream2171@ustl.edu.cn

***Lu Zhu***
Graduate School of Business and Advanced Technology Management
Assumption University
Ramkhamhaeng 24, Hua Mak, Bangkapi, Bangkok, Thailand, 10240
Tel: +6606-2098-8069
Email: zhulu@au.edu

***Changhan Li***
Graduate School of Business and Advanced Technology Management
Assumption University
Ramkhamhaeng 24, Hua Mak, Bangkapi, Bangkok, Thailand, 10240
Tel: +6606-1093-6866
Email: lichanghan@au.edu

***Abstract:*** The aim of the present study is to examine how the Chaoxing-based flipped classroom (FC) model influences Chinese English majors' critical thinking (CT) skills. To achieve this objective, this study employed a mixed-methods approach. Specifically, it integrated quantitative data from the California Critical Thinking Skills Test (CCTST) with qualitative insights from open-ended questionnaires. Based on the Solomon Four-Group Design, the study recruited 80 English majors and divided them into control and experimental groups. ANCOVA results revealed that the Chaoxing-based FC model significantly enhanced overall CT skills ($p < 0.001$, $\eta^2 = 0.234$), analysis skills ($p = 0.000$, $\eta^2 = 0.317$), and inference skills ($p = 0.001$, $\eta^2 = 0.193$). However, evaluation, interpretation, and explanation sub-skills were not significantly improved. Complementary findings obtained from qualitative analysis suggested that communication, logical reasoning, and collaboration were improved. This study also identified challenges in the cultivation of interpretation, evaluation and explanation skills, which required more targeted instructional strategies. This study highlights that digital platforms such as Chaoxing can potentially contribute to the advancement of CT and offers practical implications for how to improve FC pedagogy.

***Keywords:*** Chaoxing platform, flipped classroom, critical thinking, English majors, mixed-methods research

*Received: February 22, 2025 Revised: May 20, 2025 Accepted: June 19, 2025*

## Introduction

In the contemporary era, Critical Thinking (CT) is acknowledged as an important skill and ability that individuals should acquire for information analysis, problem solving and dynamic adaptation (Saleh, 2019). In the context of English majors in China, national policies such as the *Guidelines for English Major Education* (Trial Version) (Ministry of Education of the People's Republic of China, 2020) have emphasized that the cultivation of CT is essentially correlated with the training of innovative and all-rounded talents. What should be noted is that in conventional English majors' educational context, the training of analytical skills gives way to rote learning, with little chances given to students for their development of CT skills (Huang, 2010).

The Flipped Classroom (FC) model based on digital platforms such as Chaoxing can potentially bridge this gap. This model enables educators to shift knowledge acquisition to pre-class activities and stress interactive and reflective learning in class, helping students develop active participation and higher-order thinking (Buil-Fabregá et al., 2019). The Chaoxing platform, featured by flexibility and abundant resources, is expected to be an excellent channel for implementing the FC model in English education (Li, 2018).

The aim of the present research is to examine how the Chaoxing-based FC model influences the overall CT skills of English majors, as well as the five sub-skills, namely analysis, interpretation, inference, evaluation, and explanation. To achieve this objective, this study conducted quantitative analysis based on the Solomon Four-Group Design, as well as qualitative grounded in NVivo. It is expected that this study will offer implications for how to improve CT in Chinese undergraduates majoring in English.

## Literature Review

### Theory Relating to Flipped Classroom

The FC model was introduced by Bergmann and Sams (2012), and it significantly transforms the traditional teaching paradigm. Specifically, under this model, students are required to acquire knowledge in the pre-class stage and in class, they engage in interaction, discussion, collaboration, and problem-solving. The theoretical foundation of the FC model lines in constructivist learning theory, which posits two points: knowledge co-construction and active engagement. According to existing research, FC can effectively facilitate engagement, comprehension, and self-regulated learning, especially in medical disciplines and science, technology, engineering, and mathematics (STEM) (Al-Samarraie et al., 2020; Buil-Fabregá et al., 2019). Moreover, in the context of FC, students become the owner of their class, so they tend to develop greater independence and responsibility (Kurtz et al., 2014).

However, FC has rarely been implemented in language education, especially in undergraduate English major education in China. Zhai (2016) argued that FC can facilitate the integration of linguistic competencies and skills with higher-order thinking, which is crucial for the academic and professional success of English majors. Zhang and Zhang (2024) explored the integration of FC in a Chinese culture course for English majors and found that the model helped enhance students' autonomous learning and classroom engagement. Despite these benefits, several challenges limit

FC's broader application in within Chinese contexts. These include teachers' lack of training, long preparation time and students' unfamiliarity with self-directed learning.

**Information Technology and Chaoxing**

Traditional pedagogy has been overturned by information technology, as in the case of a popular educational platform in China called the Chaoxing Platform. With its rich functionalities such as resource sharing, interactive tools and course management, it contributed to the convenience of implementing FC (Li, 2018). Features such as real-time discussions, online quizzes, and multimedia resources significantly improve personalized learning and learning engagement (Zhang et al., 2025). By using the Chaoxing Platform, students obtain flexible access to learning materials, so they can engage in reviews according to their own learning pace and achieve deep learning.

The Chaoxing Platform also benefits EFL education. It solves problems like low participation and passive learning. Students can preview the learning materials and, in class, follow their instructors to develop higher-order thinking skills (Li, 2018). Zhang (2021) found that it helps students take more active roles in learning. Compared with other tools like Rain Classroom, Tencent Meeting, or DingTalk, Chaoxing platform is widely used in Chinese universities. It provides live teaching, video playback, online assignments, and exams. These features match the needs of FC. It is also the main teaching platform used at the author's university, facilitating technical support from the school. It fits well with the course structure and is easy for teachers and students to use.

However, there are still some problems. Students sometimes struggle with time management in self-paced learning. Previous research has noted that students may face confusion or overload when too many tasks are assigned online (Zhang, 2021). These problems show that teachers must provide clear guidance when using the platform.

**Theories Related to Critical Thinking**

It has been widely accepted that CT is a core competency in education. Its components, according to Facione (1990a), include five sub-skills that enable students to address complex issues and challenges and realize reasonable decision-making. The five sub-skills include analysis, interpretation, inference, evaluation, and explanation. The first is analysis, which breaks down complex notions into manageable components. Second, interpretation refers to understanding and clarifying the meaning of information. Third, inference is the process of logic-based conclusion drawing. Fourth, evaluation involves the assessment of whether the information is reliable and relevant. Finally, explanation lays stress on clear articulation of ideas and reasoning.

In the context of English as a major, CT is of significant relevance since in teaching English, students are required to experience advanced cognitive processing in tasks like argument construction, textual analysis, and cultural interpretation (Hu, 2008; Huang, 2010; Wu, 2015). What should be noted is that under traditional pedagogy, students have to prioritize rote memorization, so they are deprived of the opportunities to enhance truly important skills (Deng, 2012; Su, 2022). Some researchers have found new ways to teach these skills effectively in English majors' reading, writing and translation training (Su, 2022; Ma, 2023; Xiao, 2024). To help English majors develop skills in CT and problem-solving, it has become increasingly essential to accept CT as a core learning outcome.

### Research on CT Cultivation in FC

The FC model prioritizes active learning, cooperation, and reflective practices, which are consistent with CT development concepts. Empirical studies demonstrate that FC significantly enhances CT in disciplines like engineering, medicine, and business, where analytical and problem-solving skills are critical (Buil-Fabregá et al., 2019; Zhuang, 2016). Research also highlights FC's ability to foster peer-to-peer learning, encouraging students to challenge assumptions and refine their reasoning processes through interactive activities (Chen et al., 2018). In language education, recent research has also confirmed similar benefits. Ma (2023) found that EFL learners became more engaged in writing tasks and used deeper thinking strategies under FC instruction. Xiao (2024) emphasized that FC fosters learner autonomy and promotes cognitive processing in English reading courses.

Based on existing research, this study aims to integrate the Chaoxing-based FC model in the domain of English major education in China. In addition, this study integrates quantitative analysis of CT skill development and qualitative exploration of students' learning experiences, to reveal the influence of Chaoxing-based FC on the CT skills of English majors. The research findings will add to the existing body of research on how to innovate pedagogical practice in the context of digitalization.

## Methodology

### Research Questions

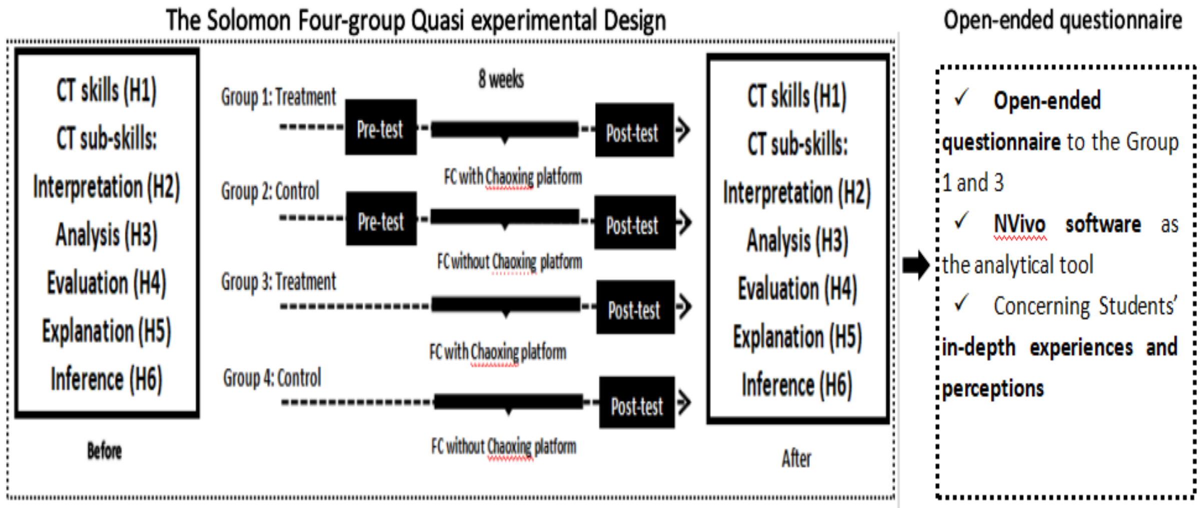
This study addresses the following research questions:

RQ1: What are the effects of the Chaoxing-based FC model on English majors' overall CT skills and its five sub-skills (interpretation, analysis, evaluation, inference, and explanation)?
RQ2: How does the Chaoxing-based FC model influence the development and improvement of CT skills among English majors?
RQ3: What are the possible factors that contribute to the observed effects?

### Conceptual Framework

The framework (Figure 1) outlines the key components, variables, and their interconnections that form the basis for this research.

**Figure 1**
*Conceptual Framework*




This study adopts a mixed-methods approach that combines a quasi-experimental Solomon Four-Group Design with qualitative data collection. The Solomon design was selected to control for potential pretest sensitization effects and improve internal validity. It allows the researcher to distinguish whether observed improvements are due to the treatment itself or influenced by exposure to the pretest. All four groups receive posttests, but Groups 1 and 2 receive both pretests and interventions, while Groups 3 and 4 only receive posttests, providing a framework to discern the unique impact of the Chaoxing-based FC model on CT skills while addressing potential confounding variables.

The intervention lasted for eight weeks and was applied to the course "Comprehensive English," which is a compulsory course for English majors. Experimental groups adopted a three-stage Chaoxing-based FC model: before class, students engaged in autonomous learning via online learning packages, such as structured video modules, quizzes, and thinking tasks on the platform; in class, they participated in discussions, debates, and peer evaluations via the classroom management functions of the Chaoxing platform, such as online discussion, group collaboration, peer review, quick response, voting, AI toolkit, etc.; after class, they completed reflective tasks and assessments on the platform. Control groups followed the same syllabus but were taught using conventional face-to-face instruction without digital tools.

Using this design helps reveal the true impact of the Chaoxing-based FC while addressing the risk of pretesting effects. However, the quasi-experimental structure may introduce ethical concerns and design biases. For instance, students in control groups may lack access to potential benefits provided by the platform. To mitigate this, the same course content was delivered to all groups, and teachers were instructed to maintain consistency in teaching objectives and assessment criteria.

Simultaneously, open-ended questionnaires collect detailed information about students' subjective experiences in Groups 1 and 3, with NVivo software serving as the analytical tool for qualitative data.

**Hypotheses**

To achieve the objectives of the study, the following hypotheses were formulated:

H1: The Chaoxing-based FC has a significant effect on English majors' overall CT skills.
H2: The Chaoxing-based FC has a significant effect on English majors' CT interpretation skill.
H3: The Chaoxing-based FC has a significant effect on English majors' CT analysis skill.
H4: The Chaoxing-based FC has a significant effect on English majors' CT evaluation skill.
H5: The Chaoxing-based FC has a significant effect on English majors' CT explanation skill.
H6: The Chaoxing-based FC has a significant effect on English majors' CT inference skill.

**Population and Sample**

To determine the sample size, this study used G*Power software, which suggested that adequate statistical power required the recruitment of 64 participants. Therefore, among the population of about 300 undergraduate English majors who were immersed in the English as a major curriculum that emphasized English language and literature in a public university in northeastern China, 80 freshmen were purposefully selected. The selection was based on class size balance, gender distribution, and consistency in course schedules. These participants represented four naturally occurring classes. They were randomly divided into four groups, each comprising 20 participants. To ensure a comparable baseline for the assessment of the influence of the intervention, the four groups were balanced in terms of previous educational background (high school graduates), age (18-23 years), and English proficiency (all had scored above 110 out of 150 on the English section of the Chinese National College Entrance Examination).

**Research Instruments**

***Performance Test: California Critical Thinking Skills Test (CCTST)***

To evaluate participants' CT skills and five sub-skills, this study employed CCTST, which is widely recognized as the "gold standard" for college-level CT assessment. Developed from the Delphi consensus framework (Facione, 1990c), the CCTST is grounded in expert-defined CT skills: interpretation, analysis, evaluation, explanation, and inference. The test contains 34 multiple-choice items that challenge participants to apply logical reasoning to scenarios presented in text, charts, or images. Each correct answer scores 1 point; incorrect responses score 0 (Facione, 1990d). Scores range from 0 to 34, with higher scores reflecting stronger CT ability. Each sub-skill has its own rating scale: "Not Manifested," "Moderate," and "Strong" (e.g., analysis: 0-2, 3-4, 5+). Given its established validity and reliability, particularly among college students, the CCTST was an appropriate measure for pretest and post-test comparisons in this investigation. To ensure participants' comprehension, the Chinese version of the CCTST was administered.

***Open-Ended Questionnaire***

This study conducted an open-ended questionnaire through the Discussion Board on the Chaoxing platform to analyze how they perceived the influence of the Chaoxing-based FC model on CT development. Questions were aligned with research objectives and designed to elicit students' reflections and experiences. This study analyzed responses using NVivo software to organize, code, and extract themes. To enhance reliability, thematic coding was first reviewed by two independent coders, and coding consistency was checked through inter-coder agreement.

***Validity and Reliability***

The CCTST, alongside expert reviews and previous research that correlated it with academic performance, as well as, confirmed strong validity (Facione, 1990b). Test-retest correlations were above 0.80, with Cronbach's alpha being 0.78, lending support to reliability (Facione, 1990b). Expert evaluation determined the content validity of the open-ended questionnaire. A Content Validity Index (CVI) of 1.00 for all items, assessed by a panel of three experts in linguistics, education, and psychology, with over 10 years of teaching experience, indicated high clarity and relevance.

**Data Analysis**

This study integrated qualitative insights from open-ended questionnaires and quantitative data collected using CCTST. Data analysis was conducted through descriptive statistics, analysis of Covariance (ANCOVA) and qualitative thematic analysis.

***Quantitative Analysis***

Participants' performance across the four Solomon Four-Group design groups was summarized by calculating descriptive statistics. Overall CT scores and five CT sub-skills were measured by metrics such as standard deviations and means.

The influence of the Chaoxing-based FC on CT skills was assessed through ANCOVA, with pretest scores controlled as covariates. The intervention (Chaoxing-based FC vs. traditional FC) served as the independent variable. Group assignment followed the Solomon Four-Group design: G1 (Pretest+Intervention), G2 (Pretest+No Intervention), G3 (No Pretest+Intervention), and G4 (No Pretest+No Intervention). All four groups completed posttests. CT total scores and five sub-skill scores were the dependent variables. Necessary assumption tests were conducted to ensure the validity of ANCOVA, and effect sizes were reported to evaluate practical significance. Post hoc pairwise comparisons with Bonferroni correction showed group differences, while partial eta squared measured the effect size.

***Qualitative Analysis***

Open-ended questionnaire responses collected via the Chaoxing platform's Discussion Board were analyzed using NVivo 12 Plus software to explore participants' reflections on the intervention. Thematic analysis was conducted through a three-stage coding process: open coding identified initial concepts, axial coding grouped these into categories, and selective coding synthesized them into core themes.

## Results

### Descriptive Statistics

Table 1 shows the post-test descriptive statistics for CT skills across the four groups. The table indicates that G1 (experimental group with pretest) outperformed other groups in most CT sub-skills. These results suggest a stronger effect of Chaoxing-based FC on fostering CT skills.

**Table 1**
*Descriptive Statistics of CT Total Score and Sub-skills Across Groups*

| Variable | Group | Post-Test | | Interpretation |
|---|---|---|---|---|
| | | Mean | S.D. | |
| **Critical Thinking Skills (DV1)** | G1 (n=20) | 18.90 | 4.983 | Strong |
| | G2 (n=20) | 15.85 | 1.927 | Moderate |
| | G3 (n=20) | 17.95 | 2.929 | Moderate |
| | G4 (n=20) | 14.05 | 3.517 | Moderate |
| **Interpretation (DV2)** | G1 (n=20) | 2.55 | .510 | Not manifested |
| | G2 (n=20) | 2.20 | .768 | Not manifested |
| | G3 (n=20) | 2.35 | .587 | Not manifested |
| | G4 (n=20) | 2.10 | .718 | Not manifested |
| **Analysis (DV3)** | G1 (n=20) | 5.00 | .795 | Strong |
| | G2 (n=20) | 4.00 | 1.170 | Moderate |
| | G3 (n=20) | 4.40 | 1.046 | Strong |
| | G4 (n=20) | 3.20 | .894 | Moderate |
| **Evaluation (DV4)** | G1 (n=20) | 2.55 | 1.468 | Not manifested |
| | G2 (n=20) | 2.20 | .951 | Not manifested |
| | G3 (n=20) | 2.40 | 1.046 | Not manifested |
| | G4 (n=20) | 2.00 | 1.338 | Not manifested |
| **Explanation (DV5)** | G1 (n=20) | 3.25 | 1.517 | Not manifested |
| | G2 (n=20) | 2.85 | 1.040 | Not manifested |
| | G3 (n=20) | 3.00 | 1.214 | Not manifested |
| | G4 (n=20) | 2.70 | 1.559 | Not manifested |
| **Inference (DV6)** | G1 (n=20) | 5.55 | 1.986 | Moderate |
| | G2 (n=20) | 4.60 | 1.314 | Not manifested |
| | G3 (n=20) | 5.80 | 1.399 | Moderate |
| | G4 (n=20) | 4.00 | 1.298 | Not manifested |

### Hypotheses Testing

ANCOVA was conducted to examine group differences in CT outcomes. Tables 2, 3, and 4 present the ANCOVA results, adjusted post-test means, and pairwise comparisons, respectively. These tables collectively demonstrate the effects of the intervention across overall CT and five sub-skills.

#### *Variables with Significant Group Differences*

The ANCOVA results showed that the Chaoxing-based FC model had a significant positive impact on overall CT skills (DV1), analysis skills (DV3), and inference skills (DV6).

As shown in Table 2, for overall CT skills (DV1), the analysis revealed a significant main effect of group ($p = 0.000$, $\eta^2 = 0.234$), indicating a large effect size. According to the report of Table 3, the adjusted Post-test means for CT skills controlling for Pre-test scores showed that experimental groups (G1: 18.893; G3: 17.972) significantly outperformed control groups (G2: 15.848; G4: 14.037). Notably, G1 had the highest adjusted post-test mean, confirming the effectiveness of the FC model. The pairwise comparisons highlighted significant differences between G1 and G2 ($p = 0.048$) and G1 and G4 ($p=0.000$), emphasizing the FC model's strong influence on enhancing students' CT overall skills in Table 4. Therefore, H1 was accepted, confirming the positive impact of the FC model on overall CT skills.

For analysis skills (DV3), as shown in Table 2, the group effect was highly significant ($p = 0.000$, $\eta^2 = 0.317$), reflecting a large effect size. G1 (5.002) and G3 (4.396) demonstrated substantially higher scores compared to G2 (4.002) and G4 (3.199) after controlling the pre-test scores in Table 3. The pairwise comparisons highlighted significant differences between G1 and G2 ($p=0.012$) and G1 and G4 ($p = 0.000$), emphasizing the FC model's strong influence on enhancing students' analytical abilities in Table 4. Therefore, H3 was accepted, confirming the significant effect of the FC model on analysis skills.

For inference skills (DV6), as shown in Table 2, the ANCOVA results also revealed significant differences between groups ($p = 0.001$, $\eta^2 = 0.193$), with G3 (5.800) achieving the highest mean, followed by G1 (5.546) in Table 3. The pairwise comparisons confirmed significant differences between G1 and G4 ($p = 0.012$) in Table 4, underscoring the model's contribution to students' inferential reasoning. Therefore, H06 was accepted.

***Variables with Non-Significant Group Differences***

In contrast, no significant group differences were observed for interpretation (DV2), evaluation (DV4), and explanation (DV5), as shown in Table 2. The ANCOVA results showed p-values exceeding 0.05 for these variables, suggesting that the FC model did not significantly impact these sub-skills. For interpretation (DV2), group means ranged from 2.10 (G4) to 2.55 (G1), but the differences were not statistically significant in Table 3. Similarly, for evaluation (DV4) and explanation (DV5), mean differences among groups were minor and did not reach statistical significance, with p-values of 0.128 and 0.142, respectively. Therefore, H2, H4 and H5 were rejected, suggesting limited effect of FC model on CT interpretation, evaluation and explanation skills.

**Table 2**

*Analysis of Covariance (ANCOVA) Results for Critical Thinking Skills Across Groups*

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared |
|---|---|---|---|---|---|---|
| DV1_pre | 5.870 | 1 | 5.870 | .471 | .495 | .006 |
| Team | 284.739 | 3 | 94.913 | 7.618 | .000 | .234 |
| DV2_pre | .594 | 1 | .594 | 1.396 | .241 | .018 |
| Team | 2.174 | 3 | .725 | 1.703 | .174 | .064 |
| DV3_pre | .374 | 1 | .374 | .381 | .539 | .005 |
| Team | 34.248 | 3 | 11.416 | 11.629 | .000 | .317 |
| DV4_pre | .126 | 1 | .126 | .084 | .773 | .001 |
| Team | 3.265 | 3 | 1.088 | .723 | .541 | .028 |
| DV5_pre | .700 | 1 | .700 | .381 | .539 | .005 |
| Team | 3.222 | 3 | 1.074 | .585 | .627 | .023 |
| DV6_pre | .787 | 1 | .787 | .335 | .564 | .004 |
| Team | 42.021 | 3 | 14.007 | 5.963 | .001 | .193 |

**Table 3**

*Adjusted Post-Test Means for Critical Thinking Skills Across Groups (Controlling for Pre-Test Scores)*

| | | Estimates | | | |
|---|---|---|---|---|---|
| Variable | Team | Mean | Std. Error | 95% Confidence Interval | |
| | | | | Lower Bound | Upper Bound |
| DV1_post | 1 | 18.893[a] | .789 | 17.320 | 20.465 |
| | 2 | 15.848[a] | .789 | 14.276 | 17.420 |
| | 3 | 17.972[a] | .790 | 16.399 | 19.546 |
| | 4 | 14.037[a] | .789 | 12.464 | 15.609 |
| DV2_post | 1 | 2.539[a] | .146 | 2.248 | 2.830 |
| | 2 | 2.206[a] | .146 | 1.916 | 2.497 |
| | 3 | 2.355[a] | .146 | 2.064 | 2.645 |
| | 4 | 2.100[a] | .146 | 1.809 | 2.390 |
| DV3_post | 1 | 5.002[a] | .222 | 4.561 | 5.444 |
| | 2 | 4.002[a] | .222 | 3.561 | 4.444 |
| | 3 | 4.396[a] | .222 | 3.954 | 4.837 |
| | 4 | 3.199[a] | .222 | 2.758 | 3.641 |
| DV4_post | 1 | 2.541[a] | .276 | 1.991 | 3.091 |
| | 2 | 2.210[a] | .276 | 1.659 | 2.761 |
| | 3 | 2.399[a] | .274 | 1.853 | 2.945 |
| | 4 | 2.000[a] | .274 | 1.454 | 2.547 |
| DV5_post | 1 | 3.239[a] | .304 | 2.634 | 3.844 |
| | 2 | 2.856[a] | .303 | 2.252 | 3.460 |
| | 3 | 3.010[a] | .304 | 2.405 | 3.615 |
| | 4 | 2.695[a] | .303 | 2.091 | 3.299 |
| DV6_post | 1 | 5.546[a] | .343 | 4.863 | 6.229 |

| Variable | Team | Mean | Std. Error | 95% Confidence Interval | |
|---|---|---|---|---|---|
| | | Estimates | | | |
| | | | | Lower Bound | Upper Bound |
| | 2 | 4.609[a] | .343 | 3.925 | 5.292 |
| | 3 | 5.800[a] | .343 | 5.117 | 6.483 |
| | 4 | 3.995[a] | .343 | 3.312 | 4.678 |

***Note.*** a. Covariates appearing in the model are evaluated at the following values: DV1_pre = 15.1822, DV2_pre = 2.2371 ,DV3_pre = 3.9762, DV4_pre = 1.5927, DV5_pre = 2.7712, DV6_pre = 4.6050.

**Table 4**

*Pairwise Comparisons of Post-Test Scores for Critical Thinking Skills Across Groups*

| Variable | (I) Team | (J) Team | Mean Difference (I-J) | Std. Error | Sig.[b] | 95% Confidence Interval for Difference[b] | |
|---|---|---|---|---|---|---|---|
| | | | Pairwise Comparisons | | | | |
| | | | | | | Lower Bound | Upper Bound |
| DV1_ post | 1 | 2 | 3.045* | 1.116 | .048 | .020 | 6.070 |
| | | 3 | .920 | 1.117 | 1.000 | -2.107 | 3.948 |
| | | 4 | 4.856* | 1.116 | .000 | 1.831 | 7.881 |
| | 2 | 1 | -3.045* | 1.116 | .048 | -6.070 | -.020 |
| | | 3 | -2.124 | 1.117 | .366 | -5.151 | .902 |
| | | 4 | 1.811 | 1.116 | .653 | -1.214 | 4.837 |
| | 3 | 1 | -.920 | 1.117 | 1.000 | -3.948 | 2.107 |
| | | 2 | 2.124 | 1.117 | .366 | -.902 | 5.151 |
| | | 4 | 3.936* | 1.117 | .004 | .907 | 6.964 |
| | 4 | 1 | -4.856* | 1.116 | .000 | -7.881 | -1.831 |
| | | 2 | -1.811 | 1.116 | .653 | -4.837 | 1.214 |
| | | 3 | -3.936* | 1.117 | .004 | -6.964 | -.907 |
| DV3_ post | 1 | 2 | 1.000* | .313 | .012 | .151 | 1.849 |
| | | 3 | .606 | .313 | .341 | -.243 | 1.456 |
| | | 4 | 1.803* | .313 | .000 | .954 | 2.652 |
| | 2 | 1 | -1.000* | .313 | .012 | -1.849 | -.151 |
| | | 3 | -.394 | .313 | 1.000 | -1.243 | .456 |
| | | 4 | .803 | .313 | .074 | -.046 | 1.652 |
| | 3 | 1 | -.606 | .313 | .341 | -1.456 | .243 |
| | | 2 | .394 | .313 | 1.000 | -.456 | 1.243 |
| | | 4 | 1.197* | .313 | .002 | .347 | 2.046 |
| | 4 | 1 | -1.803* | .313 | .000 | -2.652 | -.954 |
| | | 2 | -.803 | .313 | .074 | -1.652 | .046 |
| | | 3 | -1.197* | .313 | .002 | -2.046 | -.347 |
| DV6_ post | 1 | 2 | .938 | .485 | .342 | -.377 | 2.252 |
| | | 3 | -.254 | .485 | 1.000 | -1.567 | 1.060 |
| | | 4 | 1.551* | .485 | .012 | .238 | 2.865 |
| | 2 | 1 | -.938 | .485 | .342 | -2.252 | .377 |

| Pairwise Comparisons | | | | | | | |
|---|---|---|---|---|---|---|---|
| Variable | (I) Team | (J) Team | Mean Difference (I-J) | Std. Error | Sig.[b] | 95% Confidence Interval for Difference[b] | |
| | | | | | | Lower Bound | Upper Bound |
| | | 3 | -1.191 | .485 | .098 | -2.505 | .123 |
| | | 4 | .614 | .485 | 1.000 | -.701 | 1.929 |
| | 3 | 1 | .254 | .485 | 1.000 | -1.060 | 1.567 |
| | | 2 | 1.191 | .485 | .098 | -.123 | 2.505 |
| | | 4 | 1.805* | .485 | .002 | .491 | 3.119 |
| | 4 | 1 | -1.551* | .485 | .012 | -2.865 | -.238 |
| | | 2 | -.614 | .485 | 1.000 | -1.929 | .701 |
| | | 3 | -1.805* | .485 | .002 | -3.119 | -.491 |

***Note.*** Based on estimated marginal means

*. The mean difference is significant at the .05 level.

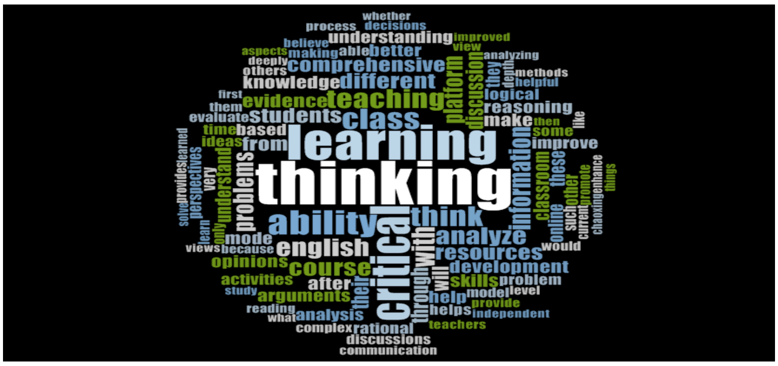
b. Adjustment for multiple comparisons: Bonferroni.

**Qualitative Results**

Within NVIVO 12 Plus, three coding methods—open, axial, and selective coding—were employed to analyze 36 open-ended questionnaires. Open coding identified key concepts and themes, axial coding refined these into categories, and selective coding synthesized them into core themes. Concepts were abstracted through repeated analysis, word cloud frequency insights, and relational synthesis, creating a structured coding hierarchy for thematic analysis（Figure 2.

**Figure 2**

*Word Cloud of Interview Materials*




During open coding, textual data from 36 responses were meticulously examined, resulting in 27 concepts and 362 reference points (Table 5). For example, students frequently highlighted enhanced analytical skills, with one participant stating: “I further analyze information or arguments by comparing different data and gathering evidence” (Student 1). Another student emphasized logical reasoning improvements: “I am adept at clearly outlining the author’s arguments and evidence, and identifying the logical chain” (Student 2), which aligns with the “strengthen logical thinking” concept under the analysis sub-skill. Axial coding consolidated these concepts into eight categories (Figure 3), such as “teaching activities preferred by students” and "the course's effect on evaluation." Finally, in selective coding, these categories were synthesized into two core themes: "The impact of FC using Chaoxing on CT" and "Evaluation of the Current Status of Teaching and Learning in FC Based on Chaoxing" (Diagram 1). After encoding with NVIVO 12 Plus, a visual project model diagram can be generated using the project model analysis tool, as illustrated in Figure 4: Project Model.

Under the first theme, participants highlighted specific CT sub-skill improvements. For inference skills, a student noted: "When confronted with controversial opinions, I can make objective judgments using acquired knowledge instead of following the crowd blindly" (Student 15), illustrating the ability to view matters rationally. On analysis skills, another participant shared: "The course taught me to deconstruct complex problems into manageable components for comprehensive analysis," (Student 12), demonstrating enhanced problem-dissection abilities.

For the second theme, students praised the platform's usability, with one stating: "Platform-based courses are flexible, allowing free allocation of study time" (Student 20), and its utility: "Chaoxing's discussion forums promote teacher-student interaction and improve learning effectiveness" (Student 21). Notably, "Flipped classrooms on the platform better enhance CT" (Student 33) emerged as a recurring sentiment, linking platform features to CT development.

**Table 5**

*Coding Examples (partial)*

| **Original Material Example** | **Open Coding** |
|---|---|
| *"I further analyze information or arguments by comparing different data and gathering evidence." (Student 1)* | Develop systematic information processing abilities |
| *"In our courses, we frequently need to scrutinize the themes, authorial perspectives, and argumentative processes of texts. I am adept at clearly outlining the author's arguments and evidence, and identifying the logical chain." (Student 2)* | Strengthen logical thinking |
| *"The course has taught me to deconstruct complex problems into smaller, more manageable components, allowing me to analyze and understand them in a more comprehensive manner." (Student 12)* | Improve the skill of dissecting complex information |
| *"When confronted with controversial opinions, I am able to employ the acquired knowledge to make objective judgments and construct arguments, rather than following the crowd blindly." (Student 15)* | View matters rationally |
| *"Platform-based course models render learning more autonomous and flexible, permitting the free allocation of study time." (Student 20)* | The schedule of platform-based courses is more flexible |
| *"Activities on the Chaoxing learning platform promote communication and interaction between teachers and students, enhancing the quality and effectiveness of teaching." (Student 21)* | Chaoxing platform's classes facilitate teacher-student communication |
| *"The teaching model based on the Chaoxing platform is beneficial for the development of students' critical thinking abilities." (Student 33)* | Flipped classrooms supported by the platform better enhance critical thinking skills |

***Note.*** Due to the large number of coding contents, only the representative or special statements are selected as examples in this study.

**Figure 3**

*Hierarchical Diagram of Encoding*

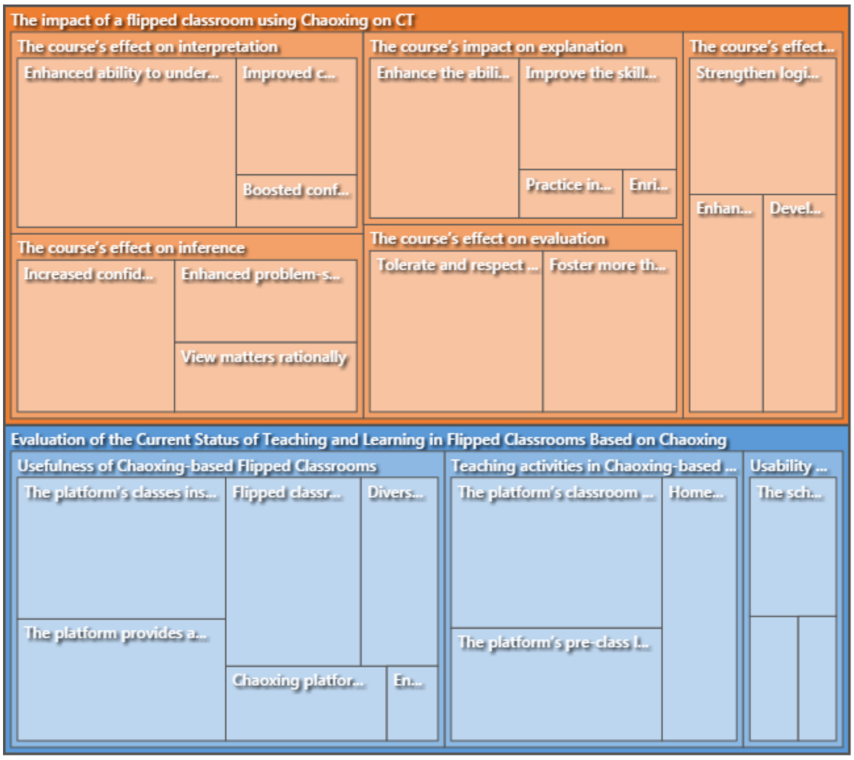


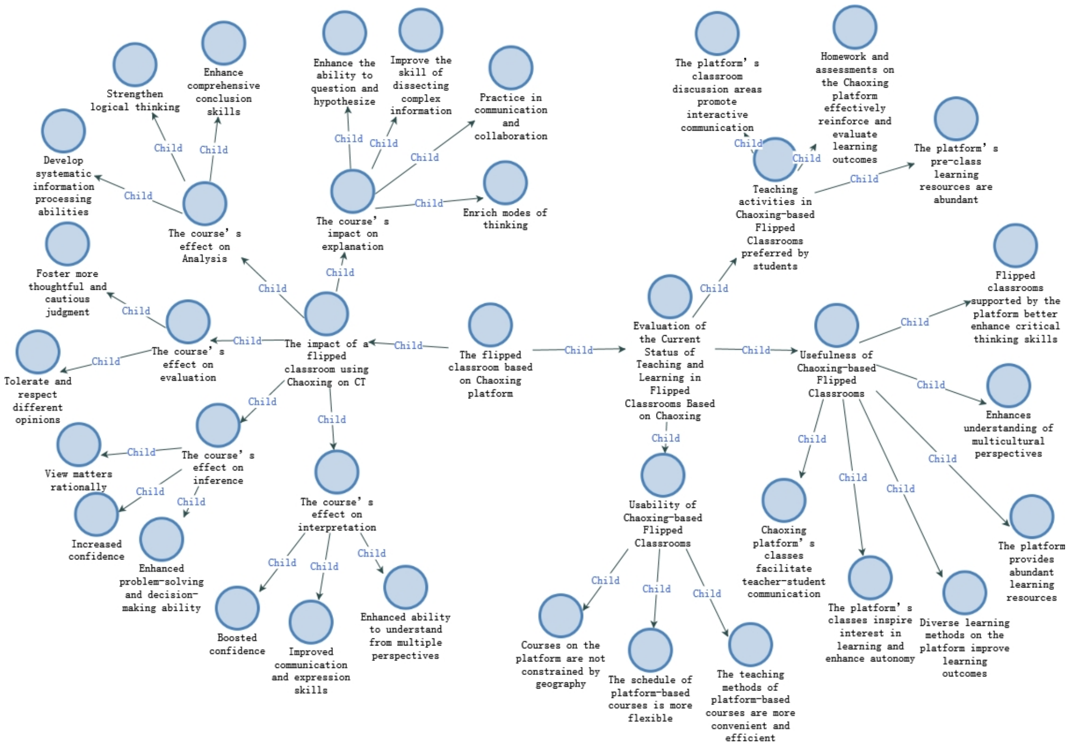

**Diagram 1**

*Tertiary Encoding Structure Diagram*

| Name | Files | References |
|---|---|---|
| Evaluation of the Current Status of Teaching and Learning in Flipped Classro | 36 | 159 |
| Teaching activities in Chaoxing-based Flipped Classrooms preferred by st | 29 | 57 |
| Homework and assessments on the Chaoxing platform effectively rei | 15 | 15 |
| The platform's classroom discussion areas promote interactive comm | 23 | 24 |
| The platform's pre-class learning resources are abundant | 18 | 18 |
| Usability of Chaoxing-based Flipped Classrooms | 15 | 19 |
| Courses on the platform are not constrained by geography | 5 | 5 |
| The schedule of platform-based courses is more flexible | 10 | 10 |
| The teaching methods of platform-based courses are more convenie | 4 | 4 |
| Usefulness of Chaoxing-based Flipped Classrooms | 30 | 83 |
| Chaoxing platform's classes facilitate teacher-student communication | 9 | 9 |
| Diverse learning methods on the platform improve learning outcome | 11 | 11 |
| Enhances understanding of multicultural perspectives | 3 | 3 |
| Flipped classrooms supported by the platform better enhance critical | 19 | 19 |
| The platform provides abundant learning resources | 15 | 19 |
| The platform's classes inspire interest in learning and enhance auton | 20 | 22 |
| The impact of a flipped classroom using Chaoxing on CT | 36 | 203 |
| The course's effect on Analysis | 24 | 39 |
| The course's effect on evaluation | 27 | 39 |
| The course's effect on inference | 31 | 41 |
| The course's effect on interpretation | 27 | 45 |
| The course's impact on explanation | 27 | 39 |

**Figure 4**

*Project Model*




Key findings were extracted through NVivo visual tools, such as word clouds and hierarchical diagrams, providing insights into recurring themes and patterns. For example, 203 reference points were identified under the theme of the flipped classroom's impact on CT, distributed across sub-skills like interpretation (45 references), inference (41 references), and evaluation (39 references) (Figure 1). Additionally, 159 references highlighted the usability, usefulness, and preferred activities in Chaoxing-based FC environments (Diagram 2). These results complement quantitative data by revealing how Chaoxing's flexible resources and collaborative activities foster active engagement, particularly in analysis and inference, while highlighting challenges in interpretation and evaluation that require targeted instructional design.

## Discussion

The research findings reveal how the Chaoxing-based FC influences English majors' CT skills. The FC model significantly influences overall CT skills (DV1), analysis skills (DV3), and inference skills (DV6), indicating that the combination of the FC model and the Chaoxing platform can contribute to the development of these skills, a finding in line with existing studies in other disciplines, demonstrating that the FC model has a positive influence on learning engagement and CT (Buil-Fabregá et al., 2019; Al-Samarraie et al., 2020). These positive changes can be attributed to reflective and interactive learning environments created by the FC model, as well as the advantages of the Chaoxing platform in terms of flexible learning and rich resources. The platform's resource-sharing capabilities before class enable comprehensive preparation and promote extensive critical engagement (DV1). Its interactive features, including discussion forums and peer assessments, strengthen analytical thinking (DV3), confirming Zhuang's (2016) view that structured pre-class tasks and peer interaction can develop students' analytical reasoning.

Furthermore, the adaptive learning environment helps monitor progress and provides personalized feedback so that students can draw evidence-based conclusions and enhance their reasoning abilities (DV6).

What should be noted is that interpretation (DV2), evaluation (DV4), and explanation (DV5) skills are not significantly improved. This echoed Dehghanzadeh and Jafaraghaee's (2018) finding that not all CT sub-skills respond equally to general flipped designs. Despite the collaborative tools and extensive resources on the platform, it may not lend full support to the development of these skills. For example, student feedback in the qualitative data indicated that interpretation (DV2) could be improved by "integrating more structured pre-class activities, such as guided annotations or in-depth content analysis exercises." Likewise, few students reported engaging in activities that required "critical assessment of arguments." Tasks that require such assessment may benefit evaluation (DV4) based on the platform's intelligent feedback features and peer review. To strengthen explanation (DV5), collaborative tasks and group discussions should involve clear articulation. As noted by Chen and Dai (2015), inference development benefits from interactive reasoning tasks, while explanation requires more deliberate guidance and feedback.

The qualitative results offer deeper insights into students' experiences, further complementing the quantitative findings. Participants said that they developed deeper understanding and greater critical engagement (DV1) during pre-class preparation, with support from the platform's structured resources, aligning with Strayer's (2012) observation that FC models improve autonomy and participation. The discussion board and collaborative in-class discussions were often shown to enhance analysis skills (DV3), since they promoted the exchange of diverse perspectives and problem deconstruction. Additionally, features such as personalized feedback and progress tracking were regarded as beneficial for inference skills (DV6). Furthermore, themes related to the effectiveness and usability of the Chaoxing-based FC environment indicates that the platform significantly promotes learning engagement and CT. However, participants also noted that it was challenging to apply higher-order thinking to tasks that involved evaluation and interpretation, suggesting that technical convenience alone is insufficient to ensure all CT skills development. Therefore, future efforts should target these gaps.

## Conclusion

This study examined the effects of the Chaoxing-based FC model on English majors' CT skills in China. The results confirmed that the model significantly improved overall CT performance, particularly in analysis and inference, while interpretation, evaluation and explanation skills showed limited progress.

Theoretically, this study extends CT research by revealing how FC combined with a digital platform influences different CT sub-skills. The differentiated outcomes underline the importance of designing skill-specific activities to target weaker areas of CT. The findings also contribute to understanding how technology-based instruction can be aligned with cognitive skill development in language learning.

Practically, this study offers implications for educators to better implement the FC model in English as a major education. Teachers should make full use of the collaborative features and structured resources on the platform to support higher-order thinking, but activities like structured debates, scaffolded reasoning tasks, and evaluation exercises may help strengthen the underdeveloped sub-skills. For policymakers, the research findings can help them design effective targeted teaching strategies characterized by the integration of digital platforms with FC approaches.

This study has limitations. First, this study only focused on freshmen of English majors from a single university, so the findings may not be generalizable to other demographic groups. Therefore, future research could include diverse groups and compare the Chaoxing-based FC model with other digital teaching methods. Secondly, there is a need for more longitudinal research on the long-term influence of FC on CT skills. Finally, to address the limited influence on interpretation, evaluation and explanation skills, it is crucial to improve the FC model by integrating targeted instructional strategies, so as to enhance the comprehensive development of CT sub-skills.

**References**

Al-Samarraie, H., Shamsuddin, A., & Alzahrani, A. I. (2020). A flipped classroom model in higher education: A review of the evidence across disciplines. *Educational Technology Research and Development, 68*(3), 1017-1051.

Bergmann, J., & Sams, A. (2012). Flip your classroom: Reach every student in every class every day. *International Society for Technology in Education, 3*(2), 71-78.

Buil-Fabregá, M., Martínez Casanovas, M., Ruiz-Munzón, N., & Filho, W. L. (2019). Flipped classroom as an active learning methodology in sustainable development curricula. *Sustainability, 11*(17), 4577.

Chen, M. H., Chao, Y. C. J., & Hung, H. T. (2018, January). Learning in a flipped English classroom from university students' perspectives. *Proceedings of the 6th International Conference on Information and Education Technology* (pp. 33-37).

Chen, X., & Dai, X. (2015). The influence of flipped classroom teaching mode on critical thinking ability of nursing undergraduates. *Huaxia Medical University, 28*(01), 142-144.

Dehghanzadeh, S., & Jafaraghaee, F. (2018). Comparing the effects of traditional lecture and flipped classroom on nursing students' critical thinking disposition: A quasi-experimental study. *Nurse education today, 71,* 151-156.

Deng, J. (2012). *A study on critical thinking disposition and critical thinking skills of English majors* [Unpublished master's thesis]. Nanchang University.

Facione, P. A. (1990a). *The California Critical Thinking Skills Test—College Level. Technical Report No. 3: Gender, ethnicity, major, CT self-esteem, and the CCTST.* California Academic Press.

Facione, P. A. (1990b). *The California Critical Thinking Skills Test—College Level. Technical Report No. 1: Experimental validation and content validity.* California Academic Press.

Facione, P. A. (1990c). *The California Critical Thinking Skills Test—College Level. Technical Report No. 2: Factors predictive of CT skills.* California Academic Press.

Facione, P. A. (1990d). *The California Critical Thinking Skills Test—College Level. Technical Report No. 4: Interpreting the CCTST, group norms, and sub-scores.* California Academic Press.

Hu, W. (2008). How professional are English majors? *Foreign Language World, 6*, 18-24.

Huang, Y. (2010). English major curriculum must be thoroughly reformed: Revisiting the issue of "lack of critical thinking". *Foreign Language World, 1*, 11–16.

Kurtz, G., Tsimerman, A., & Steiner-Lavi, O. (2014). The flipped-classroom approach: The answer to future learning. *European Journal of Open, Distance and E-Learning, 17*(2), 172-182.

Li, M. (2018). Research and practice of smart classroom teaching model based on Chaoxing learning platform. *Journal of Huaibei Vocational and Technical College, 17*(06), 36-40.

Ma, W. (2023). Cultivating critical thinking in English major writing instruction. *Journal of Social Science of Jiamusi University, 41*(2), 176–178.

Ministry of Education of the People's Republic of China. (2020). *Guidelines for English major education (Trial version)*. Higher Education Press.

Saleh, S. E. (2019). Critical thinking as a 21st century skill: Conceptions, implementation and challenges in the EFL classroom. *European Journal of Foreign Language Teaching, 4*(1), 1-16.

Strayer, J. F. (2012). How learning in an inverted classroom influences cooperation, innovation and task orientation. *Learning environments research, 15,* 171-193.

Su, X. (2022). The cultivation of critical thinking in English major translation teaching. *English Square, 8*, 76–80.

Wu, X. (2015). On the cultivation of critical thinking among English majors. *Overseas English (Part A), 5*, 100–101.

Xiao, H. (2024). Exploring approaches to cultivating critical thinking in English major reading instruction. *Modern English, 16*, 55–57.

Zhai, X. (2016). *Study on the influencing factors and mechanism of learners' satisfaction in flipped classroom* [Unpublished doctoral dissertation]. University of Science and Technology of China.

Zhang, J., & Zhang, H. (2024). A study on the application of flipped classroom in the "Introduction to Chinese Culture" course for English majors. *Qianwei,* (27), 43–45.

Zhang, L. (2021). Application of flipped classroom based on the Chaoxing learning platform in college English teaching. *Henan Agriculture, 24*, 26–27.

Zhang, X., Zhang, J., Jiang, H., Zhang, Z., & Wei, J. (2025). Practical exploration of project-based teaching of chemical engineering principles based on the Chaoxing learning platform. *Chemical Engineering Management, 4*, 43–46.

Zhuang, F. (2016). An exploration of flipped classroom mode in the cultivation of college English critical thinking. *Journal of Hubei University of Economics (Humanities and Social Sciences Edition), 13*(01), 218-220.