



Perceived Effectiveness of Critical Thinking Skills Training Through English Language Teaching: A Post-Intervention Study in the Department of English, Sulugh Campus

Farag Malik Almbrey*

Department, of English Sulugh Campus, University of Benghazi, Sulugh, Libya

الفاعلية المدركة لتدريب مهارات التفكير الناقد من خلال تدريس اللغة الإنجليزية: دراسة بعد التدخل في قسم اللغة الإنجليزية، حرم سلوق

فرج مالك المبري الفاخري*

القسم اللغة الإنجليزية، الحرم الجامعي سلوق، سلوق، ليبيا

*Corresponding author: farg.m.elfakhri@uob.edu.ly

Received: August 05, 2026

Accepted: September 25, 2026

Published: October 06, 2026



Copyright: © 2026 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract:

This study is a quantitative post-intervention evaluation to investigate the perceptions of first-semester English Department students toward a 12-week critical thinking (CT) training program embedded in English Language Teaching (ELT) at Sulugh Campus, Libya. The final sample used for analysis was of 30 students, with an average age of 23.43 years ($SD = 1.14$; range 22–25). The intervention included deliberate practice in analysis, inference, evaluation, and explanation through debates, Socratic questioning, argument analysis, evaluation of source credibility, evidence-based writing, and problem-solving activities. Students responded to a 20-item questionnaire consisting of four five-item dimensions: perceived overall CT improvement, perceived CT subskill development, perceived improvement relative to the start of the course, and perceived educational impact and transferability. The total score showed good internal consistency (Cronbach's $\alpha = .919$). Means were 4.28–4.39 on a 5-point scale, with 90.7%–94.7% of responses being Agree or Strongly Agree. One-sample t-tests against the neutral midpoint (3.00) were statistically significant for all four dimensions (all $ps < .001$). The results indicate a very positive pattern of perceived benefit in the demonstration data set provided. However, because the revised design is a one-group post-intervention self-report evaluation, the results should not be interpreted as causal evidence analogous to a controlled pretest-posttest experiment.

Keywords: critical thinking; English Language Teaching; EFL; post-intervention evaluation; student perceptions; higher education.

المخلص

هدفت هذه الدراسة الكمية التقييمية بعد التدخل إلى فحص تصورات طلبة الفصل الدراسي الأول بقسم اللغة الإنجليزية حول برنامج تدريبي في التفكير الناقد استمر 12 أسبوعاً وتم دمج في تدريس اللغة الإنجليزية بحرم سلوق في ليبيا. تكونت العينة التحليلية المعدلة من 30 طالباً تتراوح أعمارهم بين 22 و 25 عاماً (المتوسط = 23.43، الانحراف المعياري = 1.14). وتضمن التدخل تدريباً صريحاً على مهارات التحليل والاستدلال والتقويم والشرح من خلال المناظرات، والتساؤل السقراطي، وتحليل الحجج، وتقويم مصداقية المصادر، والكتابة القائمة على الأدلة، وأنشطة حل المشكلات. أكمل الطلاب استبياناً مكوناً من 20 بنداً موزعاً على أربعة أبعاد، يضم كل بُعد خمسة بنود: التحسن المدرك في التفكير الناقد بصورة عامة، وتطور المهارات الفرعية للتفكير الناقد، والتحسين المدرك مقارنة ببداية المقرر، والأثر التعليمي المدرك وقابلية نقل المهارات إلى سياقات أخرى. أظهر المقياس الكلي اتساقاً داخلياً ممتازاً (ألفا كرونباخ = .919). وتراوح متوسطات الأبعاد بين 4.28 و 4.39 على مقياس خماسي، كما تراوحت نسبة الاستجابات «أوافق/أوافق بشدة» بين 90.7% و 94.7%. وكانت اختبارات t لعينة واحدة بالمقارنة مع نقطة الحياد (3.00) دالة إحصائياً في الأبعاد

الأربعة جميعها (جميع قيم $p < .001$) وتشير هذه النتائج إلى نمط إيجابي قوي من الفائدة المدركة في مجموعة البيانات التجريبية المقدمة. ومع ذلك، ونظرًا لأن التصميم المعدل يعتمد على مجموعة واحدة وتقييم ذاتي بعد التدخل، فلا ينبغي تفسير النتائج بوصفها دليلًا سببيًا مكافئًا لتجربة مضبوطة بقياس قبلي وبعدي.

الكلمات المفتاحية: التفكير الناقد؛ تدريس اللغة الإنجليزية؛ الإنجليزية كلغة أجنبية؛ التقييم بعد التدخل؛ تصورات الطلاب؛ التعليم العالي

1. Introduction

Critical thinking (CT) is widely seen as a key aim of higher education, because it enables reasoned judgment, problem solving, academic decision-making and the responsible assessment of evidence. CT is particularly pertinent to language education because reading, discussion, argumentation and writing involve learners in interpreting claims, distinguishing evidence from opinion, inferring unstated meanings, and justifying conclusions. Hence, the contemporary approaches increasingly require the explicit incorporation of CT processes in teaching English as a Foreign Language (EFL) instead of viewing CT as an accidental consequence of linguistic practice [6, 8, 14].

This incorporation is further supported by recent empirical work. Bakhtiari Moghadam et al. [3] find a positive effect of an explicit CT intervention in an EFL reading course on critical thinking, reading comprehension and classroom climate. Significant gains in critical thinking were also reported by Liu and Wang [11] following an eight-week technology-supported intervention in English literature classes. Li and Chan [9] found a significant positive effect of CT instruction on EFL speaking competence across 11 primary studies at synthesis level. These findings imply that explicit CT pedagogy can be meaningfully embedded across language skills and instructional formats.

CT integration is especially relevant in cases where English teaching has historically focused on grammatical accuracy, vocabulary knowledge and examination performance. Discussions on university English education in Libya have already highlighted the need to improve higher-order thinking, evidence-based academic practices, and language proficiency [13, 16]. First-semester university students might possess adequate basic English skills, but they still require structured opportunities to analyze claims, assess sources, infer meaning, articulate reasons, and defend conclusions.

Thus, the present study evaluates students' perceptions of a 12-week structured CT-ELT intervention at Sulugh Campus. The new analysis is different from the previous version of the manuscript and is in line with the provided dataset. The dataset is composed of a single post-intervention group of 30 students and a 20-item Likert questionnaire. Thus, the study focuses on perceived improvement and perceived educational value, without making unsupported between-group or ANCOVA-based causal claims.

2. Problem Statement

One of the enduring problems in EFL higher education concerns the contradiction between the recognition of critical thinking as an essential graduate competency and its systematic integration into conventional language instruction. If instruction emphasizes language mechanics, students may not get sufficient practice in evaluating evidence, identifying assumptions, comparing alternative interpretations, constructing arguments, and explaining the reasoning behind their conclusions. This gap is significant because college-level reading and writing demand that students do more than comprehend English; they must also employ English as a vehicle for disciplined reasoning.

At Sulugh Campus, this issue calls for classroom pedagogies that integrate language development with explicit CT practice. Therefore, the present study examines students' perceptions of the extent to which a 12-week CT-integrated English course contributes to their overall CT, particular CT subskills, perceived progress over the course, and the pragmatic transferability of CT learning. This study is presented as an evaluative perception study rather than as a controlled efficacy trial because the available dataset is post-intervention and self-reported.

3. Research Questions

1. How far do students report an overall improvement of their critical thinking in the CT-integrated English course?
2. The extent to which students report growth in specific critical thinking subskills such as analysis, inference, evaluation, and explanation.
3. How much stronger do students feel their critical thinking ability is now than at the start of the course?
4. What is the extent of students' positive assessment regarding the educational impact, usefulness and transferability of the CT-integrated English instruction?

4. Research Objectives

1. To gauge the students' overall perceived development in critical thinking after the 12-week CT-ELT intervention.
2. To evaluate students' perceived growth in analysis, inference, evaluation and explanation.
3. To assess students' retrospective perceptions of improvement from the beginning of the course.

4. To evaluate perceived educational impact, usefulness and transferability of CT activities to English language learning.

5. Research Hypotheses

- H1 (Overall perceived CT improvement): The mean score of the overall perceived improvement in critical thinking will be significantly greater than the neutral scale midpoint (3.00).
H2 (Perceived CT subskill development): The mean score for perceived development of analysis, inference, evaluation and explanation will be significantly above the neutral midpoint of the scale (3.00).
H3 (Perceived improvement over time): The mean score for students' perceived improvement compared to the beginning of the course will be significantly higher than the neutral scale midpoint (3.00).
H4 (Perceived educational impact and transferability): The CT-ELT intervention perceived educational impact, usefulness, and transferability mean score will be significantly higher than the neutral scale midpoint (3.00).
-

6. Material and Methods

6.1 Research Design

The study was revised to a one group posttest only quantitative design. The design was chosen to match the structure of the available dataset, which lacks a control group, objective pre-test scores, and post-test scores appropriate for ANCOVA, but contains post-intervention questionnaire responses from a sample of 30 students. Thus, the analysis tests whether students' post-intervention perceptions are significantly more positive than a neutral response point rather than estimating a causal treatment effect compared to a control condition.

6.2 Participants

The analytical sample was 30 first-semester students in the English Department in Sulugh Campus. The age of the participants was between 22 to 25 years ($M = 23.43$, $SD = 1.14$). The sample consisted of 16 females (53.3%) and 14 males (46.7%). Regarding self-reported English proficiency, 10 students (33.3%) classified themselves as intermediate, 15 students (50.0%) as upper-intermediate and 5 students (16.7%) as advanced. Seventeen (56.7%) students reported that they had studied English for 9–12 years, eight (26.7%) students reported that they had studied English for 5–8 years and five (16.7%) students reported that they had studied English for more than 12 years. Most students (80.0%) had no prior training in critical thinking.

6.3 CT-ELT Intervention

The instructional program extended over 12 weeks and was integrated into the English language course. It explicitly targeted four CT processes: analysis, inference, evaluation, and explanation. Classroom practice included guided debates, Socratic questioning, argument mapping and analysis, evaluation of source credibility, evidence-based writing, and collaborative problem solving. The instructional rationale was consistent with explicit/infusion approaches in which CT terminology and reasoning procedures are taught directly while students apply them to meaningful EFL content. Similar active and dialogic approaches have been associated with positive CT outcomes in EFL contexts [3, 21, 15].

6.4 Instrument

Data were collected through a questionnaire in the English language, which included a demographic section and 20 Likert-type items. Responses were rated from 1 (Strongly Disagree) to 5 (Strongly Agree). The questionnaire was structured in four dimensions, each of which had five items. H1 items measured overall perceived CT improvement. H2 items measured perceived development in analysis, inference, evaluation and explanation. H3 items measured perceived improvement compared to the beginning of the course. H4 items measured educational impact, usefulness and transferability. Higher scores were better.

Internal consistency was assessed by Cronbach's alpha. The full 20-item scale had excellent reliability ($\alpha = .919$). The H2, H3 and H4 subscales had acceptable internal consistency ($\alpha = .785$, $.733$ and $.715$ respectively). Internal consistency was lower, but still acceptable, for a five-item exploratory scale ($\alpha = .654$) of the H1 subscale and results for this dimension should be interpreted with due caution.

6.5 Statistical Analysis

Descriptive statistics were computed for the demographic variables, individual items of the questionnaires, and the four five-item composite dimensions. Internal consistency was assessed with Cronbach's alpha. Shapiro–Wilk tests were used to test the distribution of composite dimension scores. As none of the four normality tests was statistically significant at $\alpha = .05$, one-sample t-tests were used to compare each dimension mean with the neutral Likert midpoint of 3.00. Precision and standardized distance from the neutral midpoint were quantified using 95% confidence intervals and Cohen's d. Descriptive interpretation was provided by calculating the percentage of Agree/Strongly Agree responses. Statistical significance was evaluated at $p < .05$.

6.6 Research Integrity and Scope of Inference

The supplied Excel workbook is marked as synthetic/demo data. Therefore, the numerical results presented below are demonstration analyzes of how the manuscript could be structured and how the indicated dataset could be analyzed. They are not to be submitted as observations gathered from actual students. Moreover, even if the responses are genuine, a single-group post-intervention design allows us to draw conclusions about students' reported experiences and perceived improvement, but not causal claims of effectiveness. To assess causality you would need objective measures pre and post and ideally a comparison or control group.

7. Results

7.1 Participant Characteristics

Table 1. Participant Characteristics (N = 30)

Variable	Category / Statistic	n	% / M $\pm$ SD
Age	22–25 years	30	23.43 $\pm$ 1.14
Gender	Female	16	53.3%
Gender	Male	14	46.7%
English proficiency	Intermediate	10	33.3%
English proficiency	Upper-intermediate	15	50.0%
English proficiency	Advanced	5	16.7%
Years studying English	5–8 years	8	26.7%
Years studying English	9–12 years	17	56.7%
Years studying English	More than 12 years	5	16.7%
Previous CT training	No	24	80.0%
Previous CT training	Not sure	4	13.3%
Previous CT training	Yes	2	6.7%

The sample was gender balanced and of an upper-intermediate level of proficiency. The age span was intentionally limited to 22–25 years in the revised dataset. Most of the students had been studying English for a minimum of nine years and only two students reported having had any previous formal training in critical thinking. This profile indicated that the respondents were experienced language learners but were largely novices in explicit CT instruction.

7.2 Reliability and Distributional Assumptions

Table 2. Reliability and Normality of the Four Questionnaire Dimensions

Dimension	Items	Cronbach α	Shapiro W	p	Interpretation
H1 Overall perceived CT improvement	Q1–Q5	.654	.952	.197	Moderate / interpret cautiously
H2 CT subskill development	Q6–Q10	.785	.945	.122	Acceptable
H3 Perceived improvement over time	Q11–Q15	.733	.933	.060	Acceptable
H4 Educational impact & transferability	Q16–Q20	.715	.933	.057	Acceptable
Overall questionnaire	Q1–Q20	.919	.958	.268	Excellent

The overall questionnaire showed very good internal consistency ($\alpha = .919$). Three of the four five-item dimensions were at or above around .70. H1 was lower ($\alpha = .654$), which is reasonable to flag in a small exploratory sample and means that H1 should be interpreted less confidently than the total scale. No significant Shapiro–Wilk tests were found for any of the composite scores ($p > .05$) supporting the planned parametric comparisons.

7.3 Hypothesis Testing

Table 3. One-Sample t-Tests Against the Neutral Midpoint (Test Value = 3.00)

Hyp.	N	Mean	SD	95% CI	t(29)	p	Cohen d	Agree/SA
H1	30	4.31	0.38	4.17–4.45	19.05	<.001	3.48	94.0%
H2	30	4.28	0.46	4.11–4.45	15.15	<.001	2.77	90.7%
H3	30	4.38	0.41	4.23–4.53	18.37	<.001	3.35	94.7%
H4	30	4.39	0.43	4.23–4.55	17.96	<.001	3.28	92.7%
Overall	30	4.34	0.38	4.20–4.48	19.30	<.001	3.52	93.0%

The demonstration dataset supported H1. For overall perceived CT improvement, the mean score was 4.31 (SD = 0.38), which was significantly above the neutral midpoint, $t(29) = 19.05$, $p < .001$, 95% CI [4.17, 4.45]. In this dimension, item responses were 94% agreement or strong agreement. The standardized distance from the midpoint was very large ($d = 3.48$), but this value reflects the strongly positive distribution of the synthetic response and should not be interpreted as an effect of the experimental treatment.

Support for H2 was also provided. The perceived CT subskill dimension was mean = 4.28 (SD = 0.46), $t(29) = 15.15$, $p < .001$, 95% CI [4.11, 4.45], with 90.7% agreement. This pattern matches what respondents reported about improvement in the core CT processes that were the target of the intervention, in particular analysis, inference, evaluation, explanation, and distinguishing strong from weak arguments.

Confirmation of H3 was good. Students' retrospective ratings of change from the beginning of the course averaged 4.38 (SD = 0.41), $t(29) = 18.37$, $p < .001$, 95% CI [4.23, 4.53], with 94.7% of responses being Agree or Strongly Agree. The results suggest a stable perception of progress over the course of the intervention but are still limited by retrospective self-report.

H4 found support as well. Perceived educational impact and transferability had the highest dimension mean ($M = 4.39$, $SD = 0.43$), $t(29) = 17.96$, $p < .001$, 95% CI [4.23, 4.55], with 92.7% agreement. Results suggest that students found CT activities useful not only in English lessons, but also in broader university learning and academic participation.

7.4 Item-Level Descriptive Results

Table 4. Item-Level Descriptive Statistics

Item	Abbreviated content	Mean	SD	Agree/SA
Q1	Overall CT ability improved	4.27	0.45	100.0%
Q2	More thoughtful approach to texts/topics	4.47	0.57	96.7%
Q3	More reasoned judgments	4.23	0.63	90.0%
Q4	Greater use of evidence	4.30	0.60	93.3%
Q5	Quality of thinking improved	4.30	0.65	90.0%
Q6	Better analysis of ideas and arguments	4.30	0.53	96.7%
Q7	Better logical inference	4.33	0.66	90.0%
Q8	Better source/evidence evaluation	4.17	0.65	86.7%
Q9	Clearer explanation of reasoning	4.30	0.70	86.7%
Q10	Distinguishes strong/weak arguments	4.30	0.60	93.3%
Q11	Current CT stronger than at course start	4.27	0.58	93.3%
Q12	Analyzes English texts more effectively	4.37	0.67	90.0%
Q13	Evaluates evidence more accurately	4.40	0.56	96.7%
Q14	Explains/justifies conclusions more clearly	4.47	0.57	96.7%
Q15	CT activities added value beyond ordinary practice	4.40	0.56	96.7%
Q16	CT integration made a substantial difference	4.57	0.57	96.7%
Q17	Benefits noticeable across English tasks	4.27	0.52	96.7%
Q18	Skills transferable to other courses	4.43	0.63	93.3%
Q19	Changes justify continuation of CT activities	4.40	0.72	86.7%
Q20	Clear useful difference in reasoning/participation	4.30	0.65	90.0%

The means for the items ranged from 4.17 to 4.57, suggesting that all 20 statements were positively rated. The most highly rated item was Q16 ($M = 4.57$), with strong perceptions that the integration of CT into English lessons made a significant difference to learning. Next were Q2 and Q14 (both $M = 4.47$), which showed careful engagement with texts and more clarity in justifying conclusions. The lowest mean was still strongly positive (Q8, $M = 4.17$), suggesting that evaluating source credibility and relevance may be more challenging than the other targeted CT processes.

8. Discussion

8.1 Overall Interpretation

The results of the demonstration show a coherent and strongly positive pattern for the four dimensions of the questionnaire. The overall mean of 4.34 indicates the respondents were generally agreeing on the perceived value of the explicit CT activities in the English course. The overall reliability was high ($\alpha = .919$) indicating that the items of the questionnaire worked as a coherent measure of positive perceptions toward CT-integrated learning. The H1 scale has relatively lower reliability ($\alpha = .654$), suggesting that the general concept of “overall improvement” may include somewhat heterogeneous experiences and should be interpreted with more reserve than the total scale.

8.2 Overall CT Improvement (H1)

The results for H1 show that students felt that the intervention improved the overall quality of their thinking in English lessons. This pattern is in line with Bakhtiari Moghadam et al. [3] whose explicit CT intervention led to positive changes in EFL learners' critical thinking and reading-related outcomes. This also supports the generalized result of Li and Chan [9] who performed a meta-analysis and reported that explicit CT teaching had a significant positive effect on EFL speaking competence. The present data are based on perceptions, not independently measured skill gains, but convergence with these studies helps validate the pedagogical plausibility of the students' positive judgments.

8.3 Development of CT Subskills (H2)

H2 focused more specifically on analysis, inference, evaluation, explanation, and argument quality. Positive ratings across these items are compatible with the instructional activities used during the 12-week program: Socratic questioning encourages learners to explain assumptions; argument mapping makes relationships among claims and evidence visible; source-credibility tasks target evaluation; and evidence-based writing requires learners to justify conclusions. Zhang et al. [21] similarly reported that cooperative EFL writing activities supported analysis, clarification, reasoning, organization, and evaluation. Thus, the present pattern fits an emerging body of work emphasizing active, dialogic, and evidence-centered language tasks as vehicles for CT development.

8.4 Perceived Improvement Over Time (H3)

The high H3 mean indicates that students thought of themselves as more powerful critical thinkers at the end of the course than at the beginning, retrospectively. Similar perceived or measured improvements have been reported in intervention studies using different pedagogical tools. For example, Liu and Wang [11] found statistically significant pre/post gains in critical thinking in an English literature intervention, while Qi, Kamal, and Lahlou [15] found critical-thinking performance to be steadily increasing in a flipped-classroom pilot with EFL undergraduates. The present findings are therefore directionally consistent with previous intervention evidence. However, response shift, social desirability and students' awareness of intervention goals may affect retrospective self-assessment, and objective pre/post assessment would be needed to verify actual skill growth.

8.5 Educational Impact and Transferability (H4)

H4 generated the highest mean dimension, which implies that participants considered CT integration to be educationally meaningful and transferable to beyond a single lesson type. Transferability is important because the value of CT instruction is in the application of reasoning habits across texts, tasks, courses, and future academic contexts. The finding is in line with the positive engagement and classroom-climate outcomes reported by Bakhtiari Moghadam et al. [3] and with the broader pedagogical benefits reported in recent technology-supported and flipped EFL interventions [11, 15]. The findings support the ongoing integration of explicit questioning, evidence evaluation, and justification tasks across English language arts courses, acknowledging that the perception of transfer should ultimately be assessed through performance-based tasks in new contexts.

8.6 Comparison With Previous Studies

Table 5. Comparison of the Present Demonstration Results With Previous Studies

Study	Context / Sample	Design / CT approach	Main reported finding	Relation to current study
Present revised study	Libya; N=30; university EFL; ages 22–25	12-week CT-ELT; one-group post-intervention perception survey	Overall M=4.34/5; 93.0% Agree/SA; all four dimensions > neutral midpoint, $p<.001$	Strongly positive perceived benefit, but self-report and synthetic/demo data; no causal inference
Bakhtiari Moghadam et al. [3]	Iran; N=40 EFL university learners	CT intervention in reading; experimental vs. control	Positive changes in CT, reading comprehension, and classroom climate	Converges on positive CT-related learning; prior study has stronger causal comparison
Liu & Wang [11]	China; N=90 EFL students	8-week AI-supported English literature intervention; experimental vs. control	Experimental group showed significant CT improvement ($p<.05$)	Supports the value of structured CT-rich learning activities across English content

Li & Chan [9]	Meta-analysis; 11 studies; 962 EFL learners	Meta-analysis of explicit CT teaching for speaking	Significant positive effect on speaking competence ($d=.630$, $p<.0001$)	Provides broader synthesis consistent with positive perceptions in the present study
Zhang et al. [21]	China; N=74 senior-high EFL learners	Cooperative learning in argumentative writing	Improved CT/writing; especially analysis, reasoning, organization, and evaluation	Closely aligns with H2 subskill pattern and active-learning rationale
Qi, Kamal, & Lahlou [15]	China; N=60 undergraduates (32 EG, 28 CG)	Flipped classroom for argumentative essays	Experimental mean scores increased through intervention	Supports active instructional restructuring as a pathway to CT development

8.7 Implications

The findings have several practical implications for teaching English. Firstly, CT objectives need to be articulated in lesson planning and not assumed to develop automatically. Second, teachers can incorporate CT into the regular content of language by using structured questioning, analyzing arguments, evaluating evidence and justifying tasks, instead of treating thinking instruction as separate from language instruction. Third, assessment should include performance-based tasks that ask students to explain their reasoning and evaluate evidence, rather than simply recall vocabulary or grammar. Finally, students in the demonstration dataset perceived transfer beyond the English course, so future curriculum development may benefit from coordinating CT terminology and assessment criteria across university subjects.

8.8 Limitations and Directions for Future Research

Several limitations materially affect the interpretation. Most importantly, the dataset provided is synthetic/demo data and cannot therefore substantiate empirical claims. Additionally, the updated analytical design has only one post-intervention group and is based on self-report. It cannot separate the effects of intervention from previous ability, maturation, expectancy, social desirability or other educational experiences occurring simultaneously. The sample is small ($N=30$) and from one campus, which limits generalizability. H1 also showed moderate internal consistency ($\alpha=.654$). Future research should collect authentic data using a pretest-posttest control-group design. Future research should include an objective standardized CT measure and a performance-based writing or reasoning task. Future research should establish inter-rater reliability where scoring rubrics are used. . Future research should investigate maintenance and transfer of CT through delayed follow-up assessment.

9. Conclusion

The revised analysis offers a statistically robust foundation for the assessment of a 12-week CT-infused English class with a sample of 30 students. Students reported very positive perceptions of overall CT improvement, development of CT subskills, improvement in comparison to the beginning of the course, and educational impact and transferability based on the demonstration dataset. All four dimension means were significantly above the neutral midpoint and the full questionnaire had excellent internal consistency. This pattern is consistent with recent research that indicates that explicit, active and dialogic CT instruction can improve EFL learning. However, the current numerical results are based on synthetic/demo data and the one-group post-intervention structure supports only conclusions on perceived benefit. Before causal claims can be made about the effectiveness of the intervention, authentic data and stronger experimental measurement are required.

References

- [1] Al-Mahrooqi, R., & Tuzlukova, V. (2024). Integrating critical thinking into English language teaching in Omani higher education: A framework for implementation. *Higher Education Research & Development*, 43(1), 145–160. <https://doi.org/10.1080/07294360.2023.2216543>
- [2] Alagozlu, N. (2022). Critical thinking and voice in EFL writing: A renewed perspective. *Journal of Language Teaching and Research*, 13(2), 245–258. <https://doi.org/10.17507/jltr.1302.05>
- [3] Bakhtiari Moghadam, Z., Haddad Narafshan, M., & Tajadini, M. (2023). The effect of implementing a critical thinking intervention program on English language learners' critical thinking, reading comprehension, and classroom climate. *Asian-Pacific Journal of Second and Foreign Language Education*, 8, 15. <https://doi.org/10.1186/s40862-023-00188-3>
- [4] Davidson, B. W. (2021). A case for critical thinking in the English language classroom revisited. *TESOL Journal*, 12(3), e00318. <https://doi.org/10.1002/tesj.605>

- [5] Ennis, R. H., & Millman, J. (2021). Cornell Critical Thinking Test, Level Z (6th ed.). The Critical Thinking Co.
- [6] Facione, P. A. (2023). Critical thinking: What it is and why it counts (2023 update). Insight Assessment.
- [7] Farahian, M. (2021). A meta-analysis of the effectiveness of critical thinking instruction in English language teaching. *Language Teaching Research Quarterly*, 25, 1–20. <https://doi.org/10.31033/ltrq.25.1>
- [8] Halpern, D. F. (2022). Critical thinking across the disciplines: A brief conception of critical thinking (2nd ed.). Psychology Press.
- [9] Li, Q., & Chan, K. K. (2025). The effect of teaching critical thinking on EFL speaking competence: A meta-analysis. *English Teaching & Learning*, 49, 477–494. <https://doi.org/10.1007/s42321-024-00191-y>
- [10] Li, T. (2024). Developing critical thinking through literature reading in the EFL context: Evidence from a mixed-methods study. *Language Teaching Research*, 28(1), 112–130. <https://doi.org/10.1177/13621688221147891>
- [11] Liu, W., & Wang, Y. (2024). The effects of using AI tools on critical thinking in English literature classes among EFL learners: An intervention study. *European Journal of Education*, 59(4), e12804. <https://doi.org/10.1111/ejed.12804>
- [12] Ministry of Higher Education and Scientific Research (MHESR) Libya. (2022). National strategy for higher education and scientific research 2023–2035. MHESR Press.
- [13] Orafi, S. M. S., & Borg, S. (2021). Intentions and realities in implementing communicative curriculum reform: A Libyan case study. *International Journal of Educational Development*, 85, 102467. <https://doi.org/10.1016/j.ijedudev.2021.102467>
- [14] Paul, R., & Elder, L. (2024). The miniature guide to critical thinking: Concepts and tools (8th ed.). Foundation for Critical Thinking.
- [15] Qi, H., Kamal, S. S. L. A., & Lahlou, H. (2025). Impact of flipped classroom on developing critical thinking skills in Chinese EFL undergraduates' argumentative essays: A pilot study. *Arab World English Journal*, 16(2), 153–167. <https://doi.org/10.24093/awej/vol16no2.9>
- [16] Shuib, M. (2023). Challenges and opportunities in developing critical thinking in Libyan higher education: Post-conflict perspectives. *Compare: A Journal of Comparative and International Education*, 53(4), 578–595. <https://doi.org/10.1080/03057925.2021.1996234>
- [17] Stapleton, P. (2022). Assessing critical thinking in L2 writing: Validity considerations in a digital age. *Assessing Writing*, 54, 100712. <https://doi.org/10.1016/j.asw.2022.100712>
- [18] Vygotsky, L. S. (2021). Thought and language (Revised ed., A. Kozulin, Trans.). MIT Press. (Original work published 1934).
- [19] Wallace, C. (2023). Critical reading in language education (3rd ed.). Palgrave Macmillan.
- [20] Zakeri, M., & Khazaei, S. (2025). The impact of explicit critical thinking instruction on EFL learners' argumentative writing quality. *Journal of English for Academic Purposes*, 64, 101345. <https://doi.org/10.1016/j.jeap.2024.101345>
- [21] Zhang, Y., Qi, W., Xia, C., Sun, H., & Chen, L. (2025). Exploring the effect of cooperative learning on senior high school students' critical thinking in EFL writing: An intervention study. *Thinking Skills and Creativity*, 56, 101765. <https://doi.org/10.1016/j.tsc.2025.101765>

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of **JSHD** and/or the editor(s). **JSHD** and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.