

Critical Thinking Disposition and English Argumentative Writing Self-Efficacy Among Chinese Undergraduates: The Role of Prior Generative AI Use Experience

Ying Yi*

Wenzhou Medical University, Wenzhou 325000, China

**Author to whom correspondence should be addressed.*

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

Abstract: This cross-sectional study examined associations among critical thinking disposition (CTD), English argumentative writing self-efficacy (WSE), and prior generative artificial intelligence (GenAI) use frequency among 149 Chinese undergraduate freshmen. Participants completed a theory-guided 40-item abbreviated CTD measure and a researcher-developed English argumentative WSE questionnaire. The total CTD scale and the 11-item WSE composite demonstrated excellent internal consistency (Cronbach's $\alpha = .916$ and $.952$, respectively). The primary outcome was an 11-item WSE composite that excluded a future-oriented item with limited conceptual fit to task-specific efficacy. Total CTD was moderately and positively correlated with WSE, $r = .394$, $p < .001$. In a hierarchical multiple regression analysis, gender and Gaokao English score category accounted for 3.9% of the variance in WSE. Adding prior GenAI use frequency did not significantly improve model fit, $\Delta R^2 = .002$, F change(1, 145) = 0.279, $p = .598$. Adding total CTD significantly improved model fit, $\Delta R^2 = .150$, F change(1, 144) = 26.616, $p < .001$, and the final model explained 19.1% of the variance in WSE. After adjustment for gender, Gaokao English score category, and prior GenAI use frequency, total CTD was a significant positive correlate of WSE, $B = .701$, $SE = .136$, $\beta = .388$, $p < .001$. In an exploratory model simultaneously including the four CTD dimensions, critical thinking self-confidence was the only significant unique CTD correlate of WSE. These findings indicate that Chinese undergraduate freshmen reporting stronger CTD also tended to report greater perceived capability for English argumentative writing. Because the study was cross-sectional and relied on self-report measures, causal conclusions are not warranted.

Keywords: Critical thinking disposition; English argumentative writing self-efficacy; Generative AI use frequency; Chinese undergraduates

Online publication: October 26, 2026

1. Introduction

English argumentative writing is a central component of academic literacy. It requires students to formulate

a defensible position, evaluate and integrate evidence, organize claims coherently, and address alternative viewpoints. For learners of English as a foreign language (EFL), these rhetorical and cognitive demands must be managed while students also draw on developing English-language resources. Students' beliefs about their capability to complete such tasks are therefore consequential. English argumentative writing self-efficacy (WSE) may influence learners' willingness to engage in writing, persistence when encountering difficulty, use of revision strategies, and responses to feedback.

Critical thinking disposition (CTD) may be relevant to such beliefs. Whereas critical thinking skills concern the ability to analyze, infer, and evaluate, CTD concerns an individual's habitual willingness to engage in reasoned inquiry, seek evidence, consider alternatives, and approach problems reflectively. In argumentative writing, these dispositions may be associated with students' perceived capability to generate ideas, evaluate claims and evidence, organize an argument, and defend a position. However, relatively little research has examined the association between CTD and task-specific English argumentative WSE among Chinese undergraduates.

Generative artificial intelligence (GenAI) has become part of many students' academic and writing practices. Students may use GenAI tools for idea generation, language support, structural suggestions, or feedback-like assistance. However, a broad self-report of prior use frequency does not indicate how such tools were used, whether AI-generated content was critically evaluated, or whether use was embedded in instructional activities. Accordingly, prior GenAI use frequency was treated in this study as a background variable rather than as an indicator of a causal instructional effect.

The present cross-sectional study examined associations among CTD, English argumentative WSE, and prior GenAI use frequency among Chinese undergraduate freshmen. It also examined whether total CTD explained variance in WSE beyond gender, Gaokao English score category, and prior GenAI use frequency. By focusing on a task-specific writing-efficacy outcome, the study provides preliminary correlational evidence concerning the role of students' critical-thinking dispositions in an EFL argumentative-writing context.

The study addressed the following research questions:

- (1) Is total CTD associated with English argumentative WSE?
- (2) Does prior GenAI use frequency account for additional variance in WSE after gender and Gaokao English score category are considered?
- (3) Does total CTD account for additional variance in WSE beyond gender, Gaokao English score category, and prior GenAI use frequency?
- (4) Which CTD dimensions show unique associations with WSE when entered simultaneously?

2. Literature review

2.1. Writing self-efficacy in English argumentative writing

Self-efficacy refers to individuals' beliefs about their capability to organize and execute the actions required to attain particular performances^[1]. In writing research, writing self-efficacy concerns writers' judgments of their capability to perform specific writing-related activities successfully. It is distinct from actual writing achievement: a student may feel confident without producing high-quality writing, or may possess relevant skills while underestimating their capability. Nevertheless, writing self-efficacy is important because it has been associated with motivation, goal setting, persistence, strategy use, writing anxiety, and writing

performance ^[2-4].

A central principle in self-efficacy research is task specificity. Students' confidence may vary across writing activities, such as generating ideas, using language accurately, organizing text, integrating evidence, revising drafts, and responding to feedback. Consequently, a measure of general writing confidence may not adequately represent students' beliefs about completing a particular genre or task. The theoretical rationale for task-aligned measurement is supported by ^[5]who found that task-specific efficacy judgments were stronger predictors of performance than more general efficacy measures. In L2 writing, An and Li ^[6] likewise reported that task-specific writing anxiety and self-efficacy were distinguishable from more general L2 writing measures and operated differently across task conditions.

However, recent findings also caution against assuming that more narrowly specified measures are always empirically superior. Ye and Xu ^[7] likewise found evidence that trait-level writing self-efficacy showed relative consistency across writing tasks. Taken together, these findings suggest that task specificity has a strong theoretical basis, but its empirical advantage may depend on the writing task, learner population, outcome criterion, and the degree of correspondence between the efficacy measure and the assessed performance.

Argumentative writing is a particularly appropriate context for examining task-specific WSE. In addition to producing grammatically acceptable English, writers must formulate a clear thesis, develop reasons, evaluate and use evidence, organize logical relationships among claims, and address alternative viewpoints. These demands make argumentative writing more than a language-production task; it is also a reasoning and self-regulation task. For EFL students, confidence in argumentative writing may therefore depend not only on perceived English proficiency but also on perceived capability to engage in the cognitive processes that support argument construction. Accordingly, the present study used a task-specific WSE measure because its focal outcome was students' perceived capability to complete English argumentative writing tasks, rather than because general writing self-efficacy measures are assumed to be universally less appropriate.

2.2. Critical thinking disposition and argumentative writing self-efficacy

Critical thinking is commonly described as purposeful and self-regulatory judgment involving interpretation, analysis, evaluation, inference, and explanation ^[8]. However, the possession of critical thinking skills does not guarantee that an individual will use those skills consistently. Critical thinking disposition refers to the habitual inclination or willingness to engage in reasoned inquiry, seek evidence, consider alternatives, and reflect on one's judgments ^[9]. Thus, CTD should be distinguished from both critical thinking skill and writing performance. The present study examines whether students who report stronger dispositions toward critical thought also report greater perceived capability for English argumentative writing; it does not assess actual critical thinking performance or actual essay quality.

A social-cognitive account helps explain why CTD may be associated with WSE. According to Bandura ^[1]self-efficacy beliefs are shaped by mastery experiences, vicarious experiences, social persuasion, and interpretations of affective states. Students with stronger critical thinking dispositions may be more likely to engage actively in the cognitive and self-regulatory processes involved in argumentative writing, such as examining a question from multiple perspectives, searching for evidence, organizing claims, and revising weak reasoning. Through repeated engagement with these processes, they may accumulate experiences that support stronger beliefs in their capability to complete argumentative writing tasks. This account does not imply that CTD directly causes writing competence; rather, CTD may be associated with the forms of engagement through which writing-related efficacy beliefs are developed.

The present study focuses on four CTD dimensions: analyticity, systematicity, critical thinking self-confidence, and inquisitiveness. Analyticity involves an inclination to anticipate consequences and apply reasoned judgment to problems. Systematicity refers to an organized and orderly approach to inquiry. Critical thinking self-confidence concerns trust in one's own ability to reason effectively, whereas inquisitiveness reflects intellectual curiosity and willingness to seek information and new perspectives^[9].

These dimensions may not have identical relationships with English argumentative WSE. Analyticity and systematicity may be relevant to the planning, evaluation, and organization required in argumentative writing. Inquisitiveness may encourage students to seek information and consider competing viewpoints. Critical thinking self-confidence may be especially relevant because it most directly concerns students' perceived capability to reason through complex issues. At the same time, because critical thinking self-confidence and writing self-efficacy both involve self-reported capability-related beliefs, any unique association between them should be interpreted cautiously and may partly reflect conceptual or method overlap.

Empirical evidence concerning the relationship between critical thinking and argumentative writing remains limited and mixed. Pei et al.^[10] found no significant correlation between critical thinking skills and argumentative writing performance among English majors, although their qualitative analysis suggested that students with stronger critical thinking skills produced higher-quality essays. This study is relevant to the present research but does not directly test the same constructs: it examined critical thinking skills rather than CTD, and writing performance rather than writing self-efficacy.

Intervention studies provide related evidence that critical-thinking-oriented pedagogy may support argumentative writing development. Lu^[11] found that integrating critical thinking skills into argumentative writing instruction improved students' argumentative ability, while Qi et al.^[12] reported that a flipped-classroom approach fostered critical thinking in EFL argumentative essays. However, these intervention findings cannot establish whether students with stronger pre-existing CTD report greater task-specific English argumentative WSE. Given the limited number of direct studies, the small samples in parts of the literature, and inconsistencies in the available evidence, the association between CTD and argumentative writing beliefs among Chinese undergraduates warrants further investigation.

The present study therefore examines both total CTD and the unique associations of four CTD dimensions with English argumentative WSE. It tests associations rather than causal effects and does not infer that CTD improves actual argumentative writing quality.

2.3. Prior generative AI use frequency and writing self-efficacy

GenAI tools can support aspects of EFL writing by providing ideas, language suggestions, examples of text organization, and feedback-like responses. Such functions may reduce some immediate barriers associated with drafting and revising in a second language^[13, 14].

Recent intervention studies suggest that structured ChatGPT-supported writing activities may improve writing self-efficacy. Yang et al.^[15], for example, found that ChatGPT-supported feedback was associated with improved self-efficacy in substantive revision, discourse organization, and writing conventions. Huang and Mizumoto^[16] reported enhanced writing self-efficacy among Japanese undergraduates, and Teng^[17] also found positive self-efficacy effects in an EFL writing context. These findings indicate that pedagogically guided GenAI use may support writing-related beliefs.

However, the educational implications of GenAI use depend on more than whether students have used a tool or how frequently they have used it. Intervention studies typically involve specified tasks,

pedagogical design, and structured feedback, whereas a broad measure of prior GenAI use cannot distinguish brainstorming, language editing, structural support, source evaluation, full-text generation, or non-writing use.

Frequency of use may reflect productive support-seeking and strategic revision for some students, but it may reflect passive acceptance or overreliance for others. Bouzar et al.^[18] found that more frequent ChatGPT use predicted greater reliance on the tool. Similarly, Tusino et al.^[19] identified overreliance and reduced critical engagement as potential risks of GenAI use. Werdiningsih et al.^[20] further suggested that productive engagement with AI-generated writing support involves critical evaluation, peer consultation, and the use of primary sources. These findings indicate that the educational value of GenAI may depend less on frequency alone than on use purpose, quality of engagement, critical evaluation, and instructional guidance.

The distinction between frequency and quality of use is particularly important for argumentative writing. Productive use may involve asking the tool to generate counterarguments, checking the adequacy of evidence, comparing alternative structures, or revising one's own draft after critically evaluating suggestions. Less productive use may involve accepting generated arguments without verification or delegating idea development to the tool. Concerns about accuracy, bias, authorship, privacy, and academic integrity further underscore the need for students to evaluate AI-generated output critically^[21, 22].

Because the present study assessed only broad self-reported prior GenAI use frequency, it could not distinguish between these purposes, qualities, or instructional contexts of use. It also did not measure actual writing performance or manipulate GenAI-supported instruction. Prior GenAI use frequency was therefore included as a background covariate rather than as an indicator of a causal educational effect. This cautious operationalization allows the study to test whether broad prior experience explains variance in WSE while avoiding unsupported claims about the effects of GenAI on students' writing beliefs. Accordingly, a non-significant association should not be interpreted as evidence that GenAI-supported writing instruction has no educational value; it would indicate only that this broad measure of prior use did not explain additional variance in WSE in the present sample.

3. Method

3.1. Participants

Participants were 149 Chinese undergraduate freshmen recruited from Wenzhou Medical University in Zhejiang Province, mainland China. Data were collected between September, 2025 and December, 2025 through an anonymous online questionnaire distributed via wx.cn. The sample comprised 69 men (46.3%) and 80 women (53.7%). Participation was voluntary, and participants provided informed consent before completing the questionnaire. No personally identifiable information was collected. No incentive was offered. Of 160 submitted questionnaires, 11 were excluded because of incomplete or invalid responses, resulting in the final analytic sample of 149 participants.

The study protocol was approved by the Ethics Committee of Wenzhou Medical University. Gaokao English score categories were below 110 ($n = 2$, 1.3%), 110–119 ($n = 7$, 4.7%), 120–129 ($n = 94$, 63.1%), 130–139 ($n = 45$, 30.2%), and 140 or above ($n = 1$, 0.7%). Regarding prior GenAI use frequency, five students (3.4%) reported never using AI chat or writing tools, 76 (51.0%) reported occasional use, and 68 (45.6%) reported frequent use.

3.2. Measures

3.2.1. Background variables

Participants reported their gender, Gaokao English score category, and prior frequency of using AI chat or writing tools. Gender was coded as 0 = male and 1 = female. Gaokao English score category was coded from 1 = below 110 to 5 = 140 or above. Prior GenAI use frequency was assessed with one self-report item: “Before participating in this study, how frequently had you used AI chat or writing tools (e.g., ChatGPT)?” Response options were “never used,” “tried a few times occasionally,” and “used frequently.” Responses were coded as 0, 1, and 2, respectively. Given that this single-item measure captured only broad self-reported frequency, it was treated as an exploratory background covariate rather than as an indicator of the quality, purpose, or educational effects of GenAI use.

3.2.2. Critical thinking disposition

Critical thinking disposition (CTD) was assessed using a researcher-selected 40-item shortened version of the Chinese Version of the California Critical Thinking Disposition Inventory (CCTDI-CV). The CCTDI-CV was translated and psychometrically examined by Yeh ^[23] in a sample of Taiwanese undergraduate nursing students. The original study reported an overall Cronbach’s α of .71, subscale alphas ranging from .34 to .73, and evidence of content and construct validity. Because the original validation sample and context differ from the present sample of mainland Chinese undergraduates, its psychometric evidence should not be assumed to transfer automatically to the present study.

The full CCTDI-CV contains 70 items. To reduce respondent burden and minimize potential fatigue in the online survey, the present study retained four dimensions considered most directly relevant to argumentative reasoning and writing self-efficacy: analyticity, systematicity, critical thinking self-confidence, and inquisitiveness. The shortened scale included 10 items for each dimension. Analyticity was assessed by Items 1–10, systematicity by Items 11–20, critical thinking self-confidence by Items 21–30, and inquisitiveness by Items 31–40. The four retained subscales were included in their original 10-item form and were selected a priori on theoretical grounds rather than on the basis of item statistics from the present sample.

All items were rated on a 6-point Likert-type scale ranging from 1 (strongly disagree) to 6 (strongly agree). Negatively worded items (Items 8–10, 15–20, 30, and 38–40) were reverse scored before scale scores were calculated. Mean scores were computed for the total CTD scale and each subscale, with higher scores indicating a stronger critical thinking disposition.

Because the 40-item measure was a theory-guided, researcher-selected subset of the CCTDI-CV, its internal consistency was re-examined in the present sample. The total CTD scale demonstrated excellent internal consistency (Cronbach’s $\alpha = .916$). Internal consistency was acceptable for analyticity ($\alpha = .747$) and systematicity ($\alpha = .739$), and good for critical thinking self-confidence ($\alpha = .827$) and inquisitiveness ($\alpha = .811$).

These reliability estimates support the internal consistency of the shortened scale scores in the present sample. However, internal consistency alone does not establish the intended four-factor structure, content coverage, or structural validity of a shortened measure. The results should therefore be interpreted as evidence based on a theory-guided, researcher-selected abbreviated measure rather than as a full psychometric validation of a new 40-item CTD instrument. Future research should examine the scale’s factor structure and validity using an independent and larger sample.

Item-level analyses indicated that several reverse-worded items had relatively low corrected item–total correlations. This pattern may reflect wording-related method effects and reinforces the need to interpret the abbreviated CTD scores cautiously. The dimension-level analyses were therefore treated as exploratory.

3.2.3. English argumentative writing self-efficacy

English argumentative writing self-efficacy was assessed using 12 six-point confidence items (1 = no confidence at all, 6 = complete confidence). The items assessed students' confidence in grammar, vocabulary, spelling, punctuation, idea generation, argument organization, thesis formulation, evidence use, transitions, meeting teacher requirements, obtaining good grades, and continued improvement in English writing.

The 12-item score was initially calculated as the mean of all items. However, Item 52 ("Continue to improve in English writing") assessed anticipated future improvement rather than perceived current capability to perform a defined argumentative-writing task. Because of its limited conceptual fit with task-specific self-efficacy and its comparatively low corrected item-total correlation, the primary outcome for the correlational and regression analyses was an 11-item WSE composite calculated as the mean of Items 41–51. Higher scores indicated stronger perceived self-efficacy for English argumentative writing. The 11-item WSE composite demonstrated excellent internal consistency in the present sample (Cronbach's $\alpha = .952$). The original 12-item scale also demonstrated excellent internal consistency (Cronbach's $\alpha = .946$).

An exploratory factor analysis using principal axis factoring with direct oblimin rotation was conducted as an initial examination of the internal structure of the researcher-developed WSE measure. Sampling adequacy was high, KMO = .917, and Bartlett's test of sphericity was significant, $\chi^2(66) = 1595.74$, $p < .001$. Although two initial eigenvalues slightly exceeded 1.00 (7.687 and 1.051), parallel analysis supported retention of only one factor. Accordingly, the results were interpreted as providing preliminary support for treating the questionnaire as broadly unidimensional, while recognizing that Item 52 showed weaker conceptual and empirical alignment with the remaining task-specific efficacy items.

For descriptive purposes only, inspection of a forced two-factor oblique solution suggested that Items 41–44 tended to align more closely with confidence in language conventions, whereas Items 45–49 tended to align more closely with confidence in argumentative writing processes. However, Items 50 and 51 showed cross-loadings, and Item 52 loaded comparatively weakly. In addition, any potential subdimensions were highly related. Therefore, the 11-item WSE composite excluding Item 52 was retained as the primary outcome variable for the main analyses. The 12-item score was examined only as a sensitivity analysis, and the EFA findings were treated as exploratory rather than confirmatory evidence.

3.3. Procedure

This study used a cross-sectional online survey. Participants completed an anonymous questionnaire containing demographic questions, the CTD questionnaire, and the English argumentative WSE questionnaire. Before beginning the survey, participants were informed of the study purpose and provided informed consent. The questionnaire was completed in a single sitting.

3.4. Data analysis

Descriptive statistics and Cronbach's α coefficients were calculated for the CTD scale and the original 12-item WSE questionnaire. Pearson correlation analyses were conducted to examine associations between CTD and the primary 11-item WSE composite. An EFA was conducted to examine the internal structure of the researcher-developed 12-item WSE questionnaire.

Hierarchical multiple regression analysis was conducted with the 11-item WSE composite as the dependent variable. Gender and Gaokao English score category were entered in Step 1, prior GenAI use

frequency was entered in Step 2, and total CTD was entered in Step 3. Unstandardized coefficients, standard errors, standardized coefficients, 95% confidence intervals, model fit statistics, and changes in R^2 were reported. All statistical tests were two-tailed, with $\alpha = .05$.

Visual inspection of residual plots revealed no major departures from linearity or homoscedasticity. No influential observations were identified on the basis of Cook's distance. In the exploratory CTD-dimension model, variance inflation factors ranged from 1.02 to 2.17, indicating no substantial multicollinearity concern.

4. Results

4.1. Descriptive statistics, reliability, and bivariate associations

Table 1 presents descriptive statistics and internal consistency reliability estimates for the CTD scale, its four dimensions, and the original 12-item WSE questionnaire. The total CTD scale and its four dimensions demonstrated acceptable to excellent internal consistency. The original 12-item WSE questionnaire and the 11-item WSE primary composite demonstrated excellent internal consistency. The 11-item WSE composite was used as the primary outcome in the correlational and regression analyses.

Table 1. Descriptive statistics and internal consistency reliability of the study variables

Variable	Items	<i>M</i>	<i>SD</i>	Cronbach's α
Critical thinking disposition (total)	40	4.163	.505	.916
Analyticity	10	4.351	.573	.747
Systematicity	10	3.903	.589	.739
Critical thinking self-confidence	10	4.016	.634	.827
Inquisitiveness	10	4.381	.637	.811
Writing self-efficacy, original questionnaire	12	3.651	.879	.946
Writing self-efficacy, primary composite	11	3.574	.915	.952

Note. All scale scores are item means ranging from 1 to 6. The primary WSE composite was the mean of Items 41–51; Item 52 was excluded because it assessed anticipated future improvement rather than current task-specific writing capability. WSE = English argumentative writing self-efficacy.

The total CTD score was moderately and positively correlated with the 11-item WSE composite, $r = .394$, $p < .001$. Each CTD dimension was also positively correlated with the 11-item WSE composite: analyticity, $r = .301$; systematicity, $r = .342$; critical thinking self-confidence, $r = .368$; and inquisitiveness, $r = .299$, all $ps < .001$. Thus, students reporting stronger critical thinking dispositions also tended to report greater confidence in English argumentative writing.

Table 2. Correlations between critical thinking disposition and the 11-item English argumentative writing self-efficacy composite

CTD variable	<i>r</i> with 11-item WSE
CTD total	.394***
Analyticity	.301***
Systematicity	.342***
Critical thinking self-confidence	.368***
Inquisitiveness	.299***

Note. $N = 149$. The primary WSE composite was the mean of Items 41–51. *** $p < .001$.

4.2. WSE item analysis and exploratory factor analysis

The original 12-item WSE questionnaire showed excellent internal consistency (Cronbach's $\alpha = .946$). Item 52 had the lowest corrected item–total correlation (.407), and deleting it increased Cronbach's α to .952. Because Item 52 (“Continue to improve in English writing”) concerns anticipated future improvement rather than current capability to perform a defined English argumentative-writing task, it had limited conceptual fit with the task-specific self-efficacy construct. Therefore, the primary outcome for the correlational and regression analyses was the 11-item WSE composite excluding Item 52.

The 11-item WSE composite showed excellent internal consistency (Cronbach's $\alpha = .952$). Corrected item–total correlations ranged from .659 to .861. Total CTD was positively associated with the 11-item WSE composite, $r = .394$, $p < .001$. After adjustment for gender, Gaokao English score category, and prior GenAI use frequency, total CTD remained a significant positive correlate of WSE, $B = .701$, $SE = .136$, $\beta = .388$, $t = 5.159$, $p < .001$, 95% CI [.433, .970].

Table 3. Item analysis for the primary 11-item English argumentative writing self-efficacy composite

Item	Corrected item–total correlation	Cronbach's α if item deleted
41. Correct grammar	.822	.946
42. Accurate and varied vocabulary	.811	.947
43. Correct spelling	.703	.951
44. Correct punctuation	.659	.952
45. Generate sufficient ideas	.768	.948
46. Organize ideas logically	.821	.947
47. Write a clear, strong thesis	.816	.947
48. Support claims with examples/evidence	.790	.948
49. Write smooth transitions	.804	.947
50. Meet teacher requirements	.861	.945
51. Earn a good grade	.769	.949

Note. The primary WSE composite was the mean of Items 41–51. Total-scale Cronbach's $\alpha = .952$. Item 52 was excluded because it assessed anticipated future improvement rather than current task-specific writing capability.

As shown in **Table 4**, two initial eigenvalues slightly exceeded 1.00 (7.687 and 1.051). However, parallel analysis supported retention of only one factor. Therefore, the eigenvalue-greater-than-one criterion was not treated as sufficient evidence for a two-factor solution.

For descriptive purposes only, inspection of a forced two-factor oblique pattern matrix suggested that Items 41–44 tended to align more closely with language conventions, whereas Items 45–49 tended to align more closely with argumentative writing processes. Items 50 and 51 showed substantial cross-loadings, and Item 52 loaded weakly. Because parallel analysis supported a single broad factor, the potential subdimensions were highly related, and the questionnaire was designed to assess overall perceived capability for English argumentative writing, the 11-item WSE composite excluding Item 52 was retained as the primary outcome variable. These findings should be interpreted as a preliminary exploration of item structure rather than confirmation of a multidimensional scale.

Table 4. Forced two-factor exploratory solution for the original 12-item English argumentative writing self-efficacy questionnaire

Item	Pattern coefficient: Factor 1	Pattern coefficient: Factor 2
41. Correct grammar	.474	.749
42. Accurate and varied vocabulary	.449	.766
43. Correct spelling	.321	.740
44. Correct punctuation	.349	.616
45. Generate sufficient ideas	.761	.335
46. Organize ideas logically	.838	.343
47. Write a clear, strong thesis	.805	.374
48. Support claims with examples/evidence	.797	.336
49. Write smooth transitions	.761	.387
50. Meet teacher requirements	.624	.606
51. Earn a good grade	.499	.622
52. Continue to improve in English writing	.240	.332

Note. $N = 149$. Principal axis factoring with direct oblimin rotation was used. $KMO = .917$. Bartlett's test of sphericity was significant, $\chi^2(66) = 1595.74$, $p < .001$. Although the first two initial eigenvalues were 7.687 and 1.051, parallel analysis supported retention of one factor. The forced two-factor solution is presented only to illustrate the tentative pattern of item coefficients and cross-loadings in the original 12-item questionnaire and was not retained as the primary structural interpretation. The factors were strongly correlated ($r = .787$).

4.3. Hierarchical regression predicting writing self-efficacy

Gender and Gaokao English score category entered at Step 1 explained 3.9% of the variance in the 11-item WSE composite, $R^2 = .039$, adjusted $R^2 = .026$, $F(2, 146) = 3.000$, $p = .053$. Adding prior GenAI use frequency at Step 2 did not significantly improve model fit, $\Delta R^2 = .002$, F change(1, 145) = 0.279, $p = .598$. The Step 2 model explained 4.1% of the variance in the 11-item WSE composite, $R^2 = .041$, adjusted $R^2 = .021$, $F(3, 145) = 20083$, $p = .105$.

Adding total CTD at Step 3 significantly improved model fit, $\Delta R^2 = .150$, F change(1, 144) = 26.616, $p < .001$. The final model explained 19.1% of the variance in the 11-item WSE composite, $R^2 = .191$, adjusted $R^2 = .168$, $F(4, 144) = 8.493$, $p < .001$. Total CTD was a significant positive correlate of the 11-item WSE composite, $B = .701$, $SE = .136$, $\beta = .388$, $t = 5.159$, $p < .001$, 95% CI [.433, .970]. Gender and prior GenAI use frequency were not statistically significant unique correlates at the .05 level. Gaokao English score category was a small positive unique correlate of the 11-item WSE composite, $B = .236$, $SE = .110$, $\beta = .161$, $t = 2.140$, $p = .034$.

Table 5. Hierarchical multiple regression predicting the 11-item English argumentative writing self-efficacy composite

Predictor	Step 1 <i>B</i>	<i>SE</i>	β	95% CI for <i>B</i>	Step 2 <i>B</i>	<i>SE</i>	β	95% CI for <i>B</i>	Step 3 <i>B</i>	<i>SE</i>	β	95% CI for <i>B</i>
Constant	2.922	.395	—	[2.141, 3.703]	2.825	.436	—	[1.963, 3.688]	−.100	.695	—	[−1.474, 1.274]
Gender	−.235	.149	−.128	[−.529, .060]	−.229	.150	−.125	[−.525, .067]	−.184	.138	−.101	[−.457, .089]
Gaokao English score category	.239*	.119	.164	[.004, .475]	.237*	.120	.162	[.001, .474]	.236*	.110	.161	[.018, .454]

Predictor	Step 1 <i>B</i>	<i>SE</i>	β	95% CI for <i>B</i>	Step 2 <i>B</i>	<i>SE</i>	β	95% CI for <i>B</i>	Step 3 <i>B</i>	<i>SE</i>	β	95% CI for <i>B</i>
Prior GenAI use frequency	—	—	—	—	.070	.133	.043	[−.193, .334]	.061	.123	.037	[−.182, .304]
Critical thinking disposition (total)	—	—	—	—	—	—	—	—	.701***	.136	.388	[.433, .970]
<i>R</i> ²	.039				.041				.191			
Adjusted <i>R</i> ²	.026				.021				.168			
<i>F</i> for model	3.000				2.083				8.493***			
<i>df</i> for model	(2, 146)				(3, 145)				(4, 144)			
ΔR^2	—				.002				.150			
<i>F</i> change	—				.279				26.616***			
<i>df</i> for <i>F</i> change	—				(1, 145)				(1, 144)			
<i>p</i> for <i>F</i> change	—				.598				< .001			

Note. *N* = 149. Dependent variable: the 11-item English argumentative writing self-efficacy composite (mean of Items 41–51). Gender: 0 = male, 1 = female. Gaokao category: 1–5. GenAI frequency: 0 = never, 1 = occasional, 2 = frequent.

* *p* < .05, *** *p* < .001.

4.4. Regression with CTD dimensions

In an exploratory simultaneous regression model, gender, Gaokao English score category, prior GenAI use frequency, and the four CTD dimensions were entered simultaneously. The model was significant, $R^2 = .206$, adjusted $R^2 = .166$, $F(7, 141) = 5.215$, $p < .001$. Among the CTD dimensions, critical thinking self-confidence was the only significant unique correlate of the 11-item WSE composite, $B = .362$, $SE = .158$, $\beta = .251$, $t = 2.292$, $p = .023$, 95% CI [.050, .674]. Systematicity showed a positive but non-significant association with the 11-item WSE composite, $B = .315$, $SE = .161$, $\beta = .203$, $t = 1.955$, $p = .060$. Gaokao English score category was also a significant positive background correlate, $B = .229$, $SE = .111$, $\beta = .157$, $t = 2.059$, $p = .041$. Analyticity, inquisitiveness, gender, and prior GenAI use frequency were not statistically significant unique correlates.

Variance inflation factors ranged approximately from 1.02 to 2.17, indicating no substantial multicollinearity concern.

Table 6. Exploratory simultaneous regression of the 11-item English argumentative writing self-efficacy composite on critical thinking disposition dimensions and background variables

Predictor	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>	95% CI for <i>B</i>
Constant	.041	.709	—	.057	.954	[−1.360, 1.441]
Gender	−.207	.142	−.113	−1.460	.147	[−.487, .073]
Gaokao English score category	.229	.111	.157	2.059	.041	[.009, .450]
Prior GenAI use frequency	.069	.124	.042	.553	.581	[−.177, .314]
Analyticity	.031	.173	.020	.181	.856	[−.311, .374]
Systematicity	.315	.161	.203	1.955	.053	[−.003, .633]
Critical thinking self-confidence	.362	.158	.251	2.292	.023	[.050, .674]
Inquisitiveness	−.004	.159	−.003	−.026	.979	[−.318, .310]

Note. *N* = 149. Dependent variable: 11-item English argumentative writing self-efficacy. $R^2 = .206$, adjusted $R^2 = .166$, $F(7, 141) = 5.215$, $p < .001$. Gender was coded 0 = male and 1 = female. Gaokao English score category was coded from 1 = below 110 to 5 = 140 or above. Prior GenAI use frequency was coded 0 = never, 1 = occasionally, and 2 = frequently. CTD dimension model was exploratory.

* *p* < .05

5. Discussion

This study examined the relationships among critical thinking disposition (CTD), English argumentative writing self-efficacy (WSE), and prior generative artificial intelligence (GenAI) use frequency among Chinese undergraduate students. The central finding was that overall CTD was moderately and positively associated with the 11-item WSE composite. After gender, Gaokao English score category, and prior GenAI use frequency were considered, total CTD remained a significant positive correlate of WSE and accounted for an additional 15.0% of the variance in WSE. Gaokao English score category was also a small positive unique correlate of WSE in the final model. This result should be interpreted cautiously because the score was reported in broad categories and the distribution was highly concentrated in the 120–139 range. Students who reported a stronger disposition toward reasoned, systematic, and inquisitive thinking also tended to report greater confidence in completing English argumentative writing tasks.

This finding is theoretically plausible because argumentative writing involves cognitive activities that overlap with a disposition toward critical thinking. Writers must identify issues, evaluate alternative positions, generate and assess evidence, organize claims, and articulate reasons. Students who are more willing to approach problems analytically and systematically may perceive themselves as better prepared to manage these demands. However, the cross-sectional design does not establish whether CTD contributes to WSE, whether greater writing confidence encourages critical engagement, or whether other factors, such as English proficiency, prior writing instruction, or academic achievement shape both variables.

The dimensional analysis provides an important qualification to the association between total CTD and WSE. When the four CTD dimensions were entered simultaneously, critical thinking self-confidence was the only statistically significant unique predictor of WSE. This pattern suggests that students' trust in their own ability to reason carefully may be especially relevant to confidence in undertaking English argumentative writing. Systematicity showed a marginal association with WSE, whereas analyticity and inquisitiveness did not explain unique variance after shared variance among the CTD dimensions was taken into account. This pattern does not imply that analyticity or inquisitiveness are unimportant for writing. Rather, their bivariate associations with WSE may substantially overlap with other CTD dimensions, especially critical thinking self-confidence.

Prior GenAI use frequency did not significantly predict WSE in either the background-only model or the final regression model. This finding should not be interpreted as evidence that GenAI has no educational value. The measure captured only self-reported frequency of prior use and did not distinguish between active, evaluative practices, such as critiquing AI-generated arguments, comparing evidence, and revising drafts, and passive or superficial use. In addition, the never-use group was small, limiting the precision of comparisons across use-frequency categories. Future research should assess the purpose, quality, frequency, and instructional context of GenAI use and should employ longitudinal or experimental designs when examining its potential educational effects.

The researcher-developed WSE questionnaire demonstrated excellent internal consistency. Exploratory factor analysis provided limited and preliminary evidence that some items may reflect different aspects of writing self-efficacy, such as confidence in language conventions and confidence in argumentative writing processes. However, parallel analysis supported a single broad factor; several items showed cross-loadings, and Item 52 was only weakly related to the remaining items. Item 52 assessed anticipated future improvement rather than a current judgment of capability to perform an English argumentative writing task and therefore had limited conceptual fit with task-specific self-efficacy. The primary outcome was consequently the 11-item

WSE composite excluding Item 52, which showed excellent internal consistency (Cronbach's $\alpha = .952$). The original 12-item questionnaire was retained for descriptive and exploratory measurement analyses. Further research in larger, independent samples is needed to examine the scale's internal structure and to refine its item content.

6. Implications

English argumentative writing instruction may benefit from integrating explicit critical-thinking practice with writing instruction. Activities such as evaluating the strength of claims, identifying assumptions, comparing sources, planning evidence chains, and justifying revisions may support both reasoning-related dispositions and writing confidence. Because critical thinking self-confidence showed a unique association with WSE, instructors may also provide scaffolded opportunities for students to make, explain, and revise judgments in a supportive learning environment.

In GenAI-supported writing environments, instruction should emphasize verification, critical evaluation, transparent and responsible use, and revision rather than the efficient production of text alone. GenAI may be most educationally valuable when it is positioned as an object of evaluation and a resource for revision, rather than as a substitute for students' own reasoning and composing processes.

7. Limitations and future research

Several limitations should be considered. First, the cross-sectional and correlational design precludes causal conclusions. Second, all primary constructs were measured through self-report, which may have increased associations because of common-method bias. Third, the relatively small sample of Chinese undergraduate students limits the generalizability of the findings. Fourth, Gaokao English score category and prior GenAI use frequency were measured as coarse ordinal variables, and the never-use GenAI group was particularly small.

Finally, the WSE measure was researcher-developed and requires further validation, including confirmatory factor analysis, measurement-invariance testing, and evidence of associations with external writing-performance indicators. Future studies should include scored argumentative essays, more detailed measures of English proficiency and GenAI practices, and longitudinal or intervention-based designs to clarify the temporal and potentially causal relationships among CTD, WSE, writing performance, and GenAI-supported writing practices.

8. Conclusion

Among this sample of Chinese undergraduate students, a stronger critical thinking disposition (CTD) was associated with greater English argumentative writing self-efficacy (WSE). Total CTD explained an additional 15.0% of the variance in the 11-item WSE composite beyond gender, Gaokao English score category, and prior GenAI use frequency. When the four CTD dimensions were examined simultaneously, critical thinking self-confidence was the only significant unique correlate of WSE.

Prior GenAI use frequency was not significantly associated with WSE in this cross-sectional sample. This null finding should be interpreted cautiously because the measure captured only broad self-reported use frequency rather than the purposes, quality, or instructional context of GenAI use. Overall, the findings suggest

that critical-thinking-oriented activities may be relevant to EFL argumentative writing instruction, particularly activities that help students evaluate claims and evidence, articulate judgments, and revise arguments. However, longitudinal and intervention-based research is needed before causal conclusions or instructional effects can be established.

Disclosure statement

The author declares no conflict of interest.

Funding

Wenzhou Medical University 2025 University-Level First-Class Artificial Intelligence Undergraduate Course Program (Course Name: IntelliPath for IELTS)

References

- [1] Bandura A, 1997, Self-Efficacy: The Exercise of Control. Worth Publishers.
- [2] Pajares F, 2003, Self-Efficacy Beliefs, Motivation, and Achievement in Writing: A Review of the Literature. *Reading & Writing Quarterly*, 19(2): 139–58.
- [3] Pajares M, Johnson MJ, 1993, Confidence and Competence in Writing: The Role of Self-Efficacy, Outcome Expectancy, and Apprehension. *Research in the Teaching of English*, 28.
- [4] Zimmerman BJ, Bandura A, 1994, Impact of Self-Regulatory Influences on Writing Course Attainment. *American Educational Research Journal*, 31(4): 845–62.
- [5] Pajares F, Miller MD, 1995, Mathematics Self-Efficacy and Mathematics Performances: The Need for Specificity of Assessment. *Journal of Counseling Psychology*, 42: 190–8.
- [6] An H, Li S, 2024, Task-Specific Writing Anxiety and Self-Efficacy Are Separate from General L2 Writing Anxiety and Self-Efficacy and They Have Differential Associations with the Effects of Written Corrective Feedback in Pre-Task and Within-Task Planning. *System*, 126: 103480.
- [7] Ye W, Xu Y, 2024, Modeling Writing Self-Efficacy over Time and across Tasks in an EFL Context. *System*, 121: 103242.
- [8] Facione P, 1989, Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. *Research Findings and Recommendations*, 315.
- [9] Facione NC, Facione PA, Sanchez CA, 1994, Critical Thinking Disposition as a Measure of Competent Clinical Judgment: The Development of the California Critical Thinking Disposition Inventory. *The Journal of Nursing Education*, 33(8): 345–50.
- [10] Pei Z, Zheng C, Meng Z, et al., 2017, Critical Thinking and Argumentative Writing: Inspecting the Association among EFL Learners in China. *English Language Teaching*, 10: 31–.
- [11] Chunxia L, 2021, Infusing Critical Thinking Skills into Argumentative Writing: A Study of Chinese College English Learners. *English Language and Literature Studies*, 11: 8–.
- [12] Qi H, Lin Abdullah Kamal SS, 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*.
- [13] Kasneci E, Sessler K, Küchemann S, et al., 2023, ChatGPT for Good? On Opportunities and Challenges of Large

Language Models for Education. *Learning and Individual Differences*, 103: 102274.

- [14] Kohnke L, Moorhouse BL, Zou D, 2023, ChatGPT for Language Teaching and Learning. *RELC Journal*, 54(2): 537–50.
- [15] Yang Y, Hussin S, Hashim H, 2025, Revising with Intelligence: ChatGPT Feedback and Its Impact on EFL Students' Revision and Self-Efficacy. *Forum for Linguistic Studies*, 7.
- [16] Huang J, Mizumoto A, 2024, Examining the Effect of Generative AI on Students' Motivation and Writing Self-Efficacy. *Digital Applied Linguistics*, 1: 102324.
- [17] Teng MF, 2024, "ChatGPT Is the Companion, Not Enemies": EFL Learners' Perceptions and Experiences in Using ChatGPT for Feedback in Writing. *Computers and Education: Artificial Intelligence*, 7: 100270.
- [18] Bouzar A, El Idrissi K, Ghourdou T, 2024, ChatGPT and Academic Writing Self-Efficacy: Unveiling Correlations and Technological Dependency among Postgraduate Students. *Arab World English Journal*, 1: 225–36.
- [19] Tusino T, Dewi P, Hati M, et al., 2025, ChatGPT in Academic Writing Classes: Exploring EFL Students' Perspectives. *Journal of Languages and Language Teaching*, 13: 2162–72.
- [20] Werdiningsih I, Marzuki, Indrawati I, et al., 2024, Revolutionizing EFL Writing: Unveiling the Strategic Use of ChatGPT by Indonesian Master's Students. *Cogent Education*, 11.
- [21] Cotton DRE, Cotton PA, Shipway JR, 2024, Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT. *Innovations in Education and Teaching International*, 61(2): 228–39.
- [22] Tlili A, Shehata B, Adarkwah MA, et al., 2023, What If the Devil Is My Guardian Angel: ChatGPT as a Case Study of Using Chatbots in Education. *Smart Learning Environments*, 10(1): 15.
- [23] Yeh ML, 2002, Assessing the Reliability and Validity of the Chinese Version of the California Critical Thinking Disposition Inventory. *International Journal of Nursing Studies*, 39(2): 123–32.

Publisher's note

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