Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Conversational AI as a Cognitive Scaffold or a Substitute for Thinking: A Dual-Path Model of Cognitive Offloading, Post- Response Deep Processing, and Critical Thinking

Yiding Bu
Liaoning Shihua University
Zhentai zhang
2601888596@qq.com

Liaoning Shihua University
Shuyi Wang
Liaoning Shihua University
Yuxin Li
Liaoning Shihua University

Article

Keywords: Conversational artificial intelligence, cognitive offloading, post-response deep processing, critical thinking, artificial intelligence-supported learning

Posted Date: September 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10212076/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Drawing on cognitive offloading theory and levels-of-processing theory, this study examined how conversational artificial intelligence (AI) learning support is related to university students' critical thinking. Survey data from 426 students were analyzed using partial least squares structural equation modeling and fuzzy-set qualitative comparative analysis. The results revealed two contrasting pathways. Conversational AI learning support was positively associated with both scaffolded cognitive offloading and substitutive cognitive offloading. Scaffolded cognitive offloading was positively associated with post-response deep processing, whereas substitutive cognitive offloading was negatively associated with it. Post-response deep processing was further positively associated with critical thinking. The moderating effect of AI evaluation literacy was not significant. The configurational results provided supplementary evidence that high critical thinking was more likely to occur when scaffolded cognitive offloading was present and substitutive cognitive offloading was low. These findings suggest that conversational AI may be more closely associated with critical thinking when learners continue to verify, question, reflect on, and revise AI-generated responses rather than simply accept them.

1 Introduction

Generative conversational artificial intelligence, such as ChatGPT, has rapidly entered educational learning environments and has become a common tool for students to seek explanations, feedback, and problem-solving support (Akolekar et al., 2025; N. Hu et al., 2025). Its convenient and seemingly authoritative responses may support learning, but they may also encourage answer dependence and reduce cognitive engagement (Qian, 2025; Nasr et al., 2025). This tension has renewed attention to cognitive offloading: when learners delegate information retrieval, reasoning, and judgment to artificial intelligence, do they receive meaningful support or outsource their own thinking? Previous work suggests that assistive technologies may improve immediate performance while weakening subsequent independent thinking, yet the mechanisms remain insufficiently explained (Gerlich, 2025).

Critical thinking depends on active verification, analysis, evaluation, and integration of information. In conversational AI-supported learning, what learners do after receiving generated responses is therefore crucial. When learners verify, compare, question, integrate, reflect on, and revise AI-generated responses, AI may function as a cognitive scaffold; when learners passively accept such responses, it may encourage cognitive inertia and reduce their own reasoning (Y. Hu et al., 2025; Arthi & Gandhimathi, 2025; Shafer, 2025). The central issue is therefore not only whether conversational AI is useful, but also whether learners continue to process its responses critically after receiving support.

Drawing on cognitive offloading theory and levels-of-processing theory, this study proposes a dual-path model. In the scaffolding path, conversational AI may help learners externalize basic tasks and devote more resources to verification and reflection. In the substitution path, learners may transfer judgment-related activities to AI and reduce post-response processing. The study addresses three questions: whether support quality is associated with both scaffolded and substitutive offloading; whether post-response deep processing mediates the link between offloading and critical thinking; and whether AI evaluation literacy helps explain why similar support may produce scaffolding or substitution. Methodologically, partial least squares structural equation modeling (PLS-SEM) tests net, mediating, and moderating associations, whereas fuzzy-set qualitative comparative analysis (fsQCA) identifies configurational pathways to high and low critical thinking. By distinguishing scaffolded and substitutive forms of cognitive offloading, the study explains why similar AI support may be associated with different patterns of post-response processing and critical thinking. The overall research process is shown in Fig. 1.

2 Literature Review and Theoretical Foundations

2.1 Conversational Artificial Intelligence Learning Support

Conversational AI learning support refers to the capacity of generative dialogue systems to provide explanations, feedback, follow-up questions, and task guidance (Nguyen, 2023). Unlike ordinary information retrieval tools, conversational systems participate in problem definition, evidence organization, and viewpoint generation, thereby shaping learners' allocation of cognitive resources (Körner et al., 2026; Ortega-Ochoa et al., 2024). High-quality responses, however, may not always lead to deep learning. Clear and complete responses may either scaffold thinking or trigger answer dependence (Hou et al., 2026), making it necessary to examine post-response processing rather than support quality alone.

2.2 Cognitive Offloading Theory

Cognitive offloading theory explains how individuals use external tools to reduce internal cognitive load (Guo & Ye, 2026). In AI-supported learning, offloading is not a single-risk mechanism. Scaffolded cognitive offloading refers to delegating lower-level tasks, such as information retrieval and material organization, to release resources for analysis and reflection. Substitutive cognitive offloading refers to outsourcing reasoning, evaluation, and argumentation, thereby reducing active monitoring and independent judgment (Vendrell & Johnston, 2026; Anders & Speltz, 2025; Yavich, 2025). The distinction lies not in whether learners use AI, but in whether they retain responsibility for judgment-related cognition.

2.3 Post-Response Deep Processing and Critical Thinking

Levels-of-processing theory argues that learning outcomes depend on semantic, organizational, and reflective processing (Peng et al., 2025; Luchini et al., 2024). Post-response deep processing is defined here as secondary cognitive activity after receiving AI responses, including verification, comparison, questioning, integration, reflection, and revision. It is conceptually distinct from general engagement because it focuses on whether learners re-assume responsibility for evidence evaluation and meaning construction after externally generated answers are provided. Critical thinking refers to the ability to analyze, evaluate, reason about, and self-correct information based on evidence (Loyens et al., 2023; Tan et al., 2023). In conversational AI-supported environments, it involves identifying uncertainties, comparing explanations, and revising judgments (Balart et al., 2026). Because generated responses can appear complete, critical thinking may depend more on continuous post-response processing than on support quality alone (Zhou et al., 2024; Tang & Putra, 2025).

2.4 Artificial Intelligence Evaluation Literacy

AI evaluation literacy refers to learners' ability to understand limitations of generated outputs, identify biases, and verify information (Dong et al., 2025; Kim & Ryoo, 2026). Learners with higher AI evaluation literacy may be more cautious when using generated answers and may be less likely to rely on them without verification. However, evaluative knowledge may not automatically become verification behavior (Shen, 2026; Badali et al., 2026; Yan et al., 2026). AI evaluation literacy is therefore examined as a potential boundary condition and configurational factor.

3 Research Framework and Hypotheses

3.1 Construction of the Theoretical Model

To explain when conversational AI functions as a cognitive scaffold rather than a substitute for thinking, this study constructs a dual-path model based on cognitive offloading theory and levels-of-processing theory. The basic

assumption of the model is that the quality of conversational AI learning support is not directly linked to critical thinking outcomes; rather, this relationship may depend on how learners allocate and re-assume cognitive responsibility (Zhai et al., 2024; Li et al., 2026; Gerlich, 2025). On the one hand, the quality of conversational AI learning support may be positively related to scaffolded cognitive offloading, through which lower-level tasks are externalized, allowing learners to devote more cognitive resources to post-response deep processing. On the other hand, it may also be positively related to substitutive cognitive offloading, through which learners transfer judgment, reasoning, and argumentation to AI, a process that is expected to be negatively associated with post-response deep processing (Iqbal et al., 2025; Lin et al., 2026). Therefore, post-response deep processing is positioned as the central pathway linking cognitive offloading and critical thinking, while AI evaluation literacy is incorporated into the model to examine the potential boundary role arising from differences in learners' capabilities (Wang, 2026; Yang et al., 2025).

As shown in Fig. 2, the quality of conversational AI learning support is associated with critical thinking through both a scaffolding path and a substitution path. The scaffolding path highlights AI's supportive role in basic learning tasks: scaffolded cognitive offloading may free cognitive resources for post-response deep processing and is positively associated with critical thinking. The substitution path highlights the convenience and apparent authority of AI-generated answers, which may encourage outsourcing of thinking; thus, substitutive cognitive offloading is negatively associated with post-response deep processing and unfavorable to critical thinking. The key issue is not whether AI provides answers, but whether learners continue to verify, compare, question, and revise them.

3.2 Research Hypotheses

Based on the above model, this study proposes research hypotheses from three perspectives: the scaffolding path, the substitution path, and the boundary condition. The scaffolding path focuses on how AI supports higher-order processing by releasing lower-level cognitive load. The substitution path focuses on how AI may weaken learners' active judgment by directly generating answers. The boundary condition focuses on whether AI evaluation literacy is related to the link between substitutive cognitive offloading and post-response deep processing (see Fig. 3).

High-quality conversational AI learning support may help learners handle basic tasks such as information retrieval, organization, and explanation, thereby allowing them to allocate limited cognitive resources to analysis and judgment (Du et al., 2025; Meng et al., 2025).

H1: The quality of conversational AI learning support is positively associated with scaffolded cognitive offloading.

When AI responses are overly complete, immediate, or apparently authoritative, learners may reduce their investment in independent analysis and argumentation, which may be reflected in substitutive cognitive offloading (Klingbeil et al., 2024; Visser et al., 2025).

H2

The quality of conversational AI learning support is positively associated with substitutive cognitive offloading.

By reducing lower-level cognitive load, scaffolded cognitive offloading may create conditions for learners to continue verifying, comparing, and reflecting on AI responses (Liu et al., 2025).

H3

Scaffolded cognitive offloading is positively associated with post-response deep processing.

Substitutive cognitive offloading means that learners outsource core judgment-related responsibility to AI; therefore, it may be associated with reduced engagement in deep processing of AI responses (Tian & Zhang, 2026; Stadler et al., 2024).

H4

Substitutive cognitive offloading is negatively associated with post-response deep processing.

Critical thinking depends on the continuous evaluation of evidence, logic, and differences among viewpoints. Post-response deep processing may therefore serve as an important process through which AI responses are connected with higher-order thinking (Lee et al., 2024; Wang et al., 2025).

H5

Post-response deep processing is positively associated with critical thinking.

If scaffolded cognitive offloading is positively associated with post-response deep processing, its association with critical thinking may operate through this processing mechanism (Melanou & Beege, 2026; Li et al., 2023).

H6

Scaffolded cognitive offloading has a positive indirect association with critical thinking through post-response deep processing.

If substitutive cognitive offloading is negatively associated with post-response deep processing, its negative association with critical thinking may also operate through insufficient processing (Tian & Zhang, 2025).

H7

Substitutive cognitive offloading has a negative indirect association with critical thinking through post-response deep processing.

Learners with higher AI evaluation literacy may be more likely to identify uncertainties in AI outputs and continue verifying and reflecting on AI responses when substitutive cognitive offloading occurs (Bećirović et al., 2025; Clark, 2026; Zhu et al., 2025).

H8

AI evaluation literacy may buffer the negative association between substitutive cognitive offloading and post-response deep processing.

From the perspective of the overall mechanism, AI evaluation literacy may be related to the strength of the substitution path and may help explain whether conversational AI learning support is associated with scaffolding-oriented or substitution-oriented outcomes (Zhong & Liu, 2025; Wang, 2026).

H9

AI evaluation literacy may moderate the indirect association between substitutive cognitive offloading and critical thinking through post-response deep processing.

In summary, this study develops a hypothesis system centered on the pathway from the quality of conversational AI learning support to cognitive offloading, then to post-response deep processing, and finally to critical thinking. On the one hand, the quality of conversational AI learning support may be positively associated with critical thinking through scaffolded cognitive offloading and post-response deep processing. On the other hand, it may be negatively associated with post-response deep processing and critical thinking through substitutive cognitive offloading.

4 Research Methodology

4.1 Data Collection and Measurement

The study combined contextualized learning tasks with a structured questionnaire. All core constructs were measured on a seven-point Likert scale. A pilot test with 42 university students who had experience using conversational AI for learning examined item comprehensibility, length, and contextual appropriateness. After removing repetitive items and revising unclear wording, the formal questionnaire collected demographic information and AI-use background variables.

The formal sample consisted of university learners with prior experience using conversational AI for learning. Data were collected between January 5, 2026 and April 20, 2026. Participants used a familiar mainstream conversational AI tool to complete an information-comprehension or viewpoint-analysis task and then completed the questionnaire. Responses were excluded for abnormal completion time, failed attention checks, severe missing values, or repeated submission patterns. Of 520 distributed questionnaires, 468 were returned; after screening, 426 valid responses remained, yielding an effective response rate of 81.92%. The basic sample characteristics are reported in Table 1.

Table 1
Basic Sample Characteristics (N = 426)

Code	Sample characteristic	Category	Frequency	Percentage
A1	Gender	Male	189	44.4%
A2		Female	237	55.6%
B1	Grade level	Freshman	90	21.1%
B2		Sophomore	122	28.6%
B3		Junior	126	29.6%
B4		Senior or above	88	20.7%
C1	Disciplinary background	Humanities and social sciences	112	26.3%
C2		Science	92	21.6%
C3		Engineering	145	34.0%
C4		Education, management, or others	77	18.1%
D1	Frequency of AI use	Almost every day	108	25.4%
D2		Several times per week	177	41.5%
D3		Several times per month	99	23.2%
D4		Rarely	42	9.9%

Table 1. Basic Sample Characteristics (N = 426)

Code	Sample characteristic	Category	Frequency	Percentage
E1	Duration of AI use	Less than 6 months	96	22.5%
E2		6 months to 1 year	151	35.4%
E3		1 to 2 years	123	28.9%
E4		More than 2 years	56	13.1%

Measures covered six constructs: quality of conversational AI learning support (AIQ), scaffolded cognitive offloading (SCO), substitutive cognitive offloading (SUCO), post-response deep processing (PRDP), AI evaluation literacy (AEL), and critical thinking (CT). The sources of the measurement scales are summarized in Table 2. Existing scales were adapted where available, and SCO, SUCO, and PRDP were developed or contextualized for conversational AI-supported learning. Experts in educational technology, learning sciences, and research methodology reviewed conceptual relevance, clarity, and contextual appropriateness; pilot-test feedback further refined wording. The development and content validity procedures for the self-developed constructs are summarized in Table 3. The full measurement items and scoring scheme are provided in Supplementary Material 1. The construct measurement and item design are summarized in Table 4.

Table 2
Sources of the Measurement Scales

Code	Construct	Item source	Adaptation description
A1	Quality of conversational AI learning support (AIQ)	Scales related to AI learning support, information system quality, and intelligent tutoring systems (Wang et al., 2025; Markus et al., 2025; Wang & Chuang, 2024)	Adapted to the conversational AI learning context
B1	Scaffolded cognitive offloading (SCO)	Self-developed based on cognitive offloading theory	Emphasizes lower-level task externalization and cognitive resource release
C1	Substitutive cognitive offloading (SUCO)	Self-developed based on cognitive offloading, technology dependence, and AI dependence research	Emphasizes the outsourcing of judgment, reasoning, and argumentation
D1	Post-response deep processing (PRDP)	Self-developed based on levels-of-processing theory, self-explanation, and reflective learning research	Emphasizes verification, comparison, questioning, integration, and revision
E1	AI evaluation literacy (AEL)	Scales related to AI literacy and GenAI literacy (Ng et al., 2024; Lintner, 2024; Yuan et al., 2024)	Emphasizes error identification and output verification

Table 2. Sources of the Measurement Scales

Code	Construct	Item source	Adaptation description
F1	Critical thinking (CT)	Critical thinking disposition or skill scales (Rodríguez-Rojas et al., 2024; Li & Liu, 2024)	Adapted to learning-task contexts

Table 3
Development and Content Validity of Self-Developed Construct Items

Code	Construct	Theoretical basis	Focus of item development	Content validity procedure
A1	SCO	Cognitive offloading theory and cognitive load research	Emphasizes the use of AI to assist with lower-level tasks, such as information retrieval, material organization, and initial explanation, thereby releasing resources for higher-order thinking	Based on expert feedback, expressions related to "lower-level task externalization" and "cognitive resource release" were adjusted, and items overlapping with general AI use frequency were removed
B1	SUCO	Cognitive offloading theory, technology dependence, and AI dependence research	Emphasizes the outsourcing of higher-order cognitive responsibilities, such as judgment, reasoning, argumentation, and conclusion formation, to AI	Based on expert feedback, "tool assistance" was distinguished from "thinking substitution" to avoid confusion with scaffolded cognitive offloading
C1	PRDP	Levels-of-processing theory, self-explanation, and reflective learning research	Emphasizes verification, comparison, questioning, integration, reflection, and revision after receiving AI responses	Based on pilot-test feedback, item wording was optimized to focus on secondary processing behaviors "after AI responses"

Table 4
Construct Measurement and Item Design

Code	Construct	Number of items	Measurement content	Example item
A1	AIQ	5	Clarity of explanation, timeliness of feedback, follow-up support, and interaction effectiveness	Conversational AI can provide clear explanations for my learning questions.
B1	SCO	4	Basic task externalization, cognitive resource release, and support for higher-order processing	I use AI to handle basic tasks such as organizing materials so that I can devote more effort to analysis and judgment.
C1	SUCO	4	Judgment outsourcing, reasoning outsourcing, answer dependence, and reduced independent thinking	After AI provides an answer, I usually no longer further judge whether it is reasonable.
D1	PRDP	6	Verification, comparison, questioning, integration, reflection, and revision	After receiving an AI response, I further verify whether its viewpoints or evidence are reliable.
E1	AEL	5	Error identification, bias judgment, credibility evaluation, and information verification	I can identify possible errors, biases, or uncertainties in AI responses.

Table 4. Construct Measurement and Item Design

Code	Construct	Number of items	Measurement content	Example item
F1	CT	6	Evidence evaluation, logical analysis, viewpoint comparison, and reflective revision	When facing different viewpoints, I compare their evidence and logical basis.

Procedural controls for common method bias included anonymous responses, randomized item presentation, attention checks, and instructions that there were no right or wrong answers. Harman's single-factor test and full collinearity VIF diagnostics were also conducted.

Ethical approval

for this study was obtained from the Ethics Committee of the School of Art and Design, Liaoning Petrochemical University (Approval No. LNSH-2025010516). All procedures involving human participants were performed in accordance with the ethical standards of the approving committee, relevant institutional and national guidelines and regulations, and the Declaration of Helsinki. Electronic informed consent was obtained from all participants before data collection. Participation was voluntary, and no directly identifiable personal information was collected. All data were anonymized and used solely for academic research.

4.2 Data Analysis

PLS-SEM and fsQCA were used complementarily. PLS-SEM examined net associations, explanatory power, direct paths, mediation, and moderation (Richter & Tudoran, 2024). SmartPLS 4 was used with reflective measurement models and 5,000 bootstrap resamples. Reliability and validity were evaluated using outer loadings, Cronbach's alpha, composite reliability, average variance extracted, and HTMT; structural paths were evaluated using β, t, p, f^2, R^2, and adjusted R^2. fsQCA identified configurations linked to high and low critical thinking and captured asymmetric patterns not directly represented in linear models (Grunwald et al., 2025). The data analysis procedures and judgment criteria are summarized in Table 5.

Table 5
Data Analysis Procedures and Judgment Criteria

Code	Analytical stage	Analytical content	Judgment criteria
A1	Data preprocessing	Screening of missing values, outliers, and invalid questionnaires	Responses with abnormal completion times, failed attention checks, and severe missing data were excluded
B1	Common method bias test	Harman's single-factor test and VIF diagnostics	The variance explained by a single factor should not be excessively high, and VIF should be below the critical value
C1	Measurement model assessment	Reliability, convergent validity, and discriminant validity	$\alpha > 0.70$, CR > 0.70, AVE > 0.50, HTMT < 0.90
D1	Structural model assessment	Direct associations, explanatory power, and effect size	Path significance, R^2, adjusted R^2, and f^2 were assessed and interpreted according to the theoretical model
E1	Mediation effect test	Indirect effects of the scaffolding path and substitution path	Bootstrap confidence intervals should not include 0
F1	Moderation effect test	Boundary role of AI evaluation literacy	A significant interaction term indicates support for the moderation hypothesis
G1	fsQCA calibration	Transformation of latent variable scores into fuzzy sets	Full membership, crossover point, and full non-membership thresholds were set
H1	Necessary condition analysis	Determining whether a single condition is necessary	Consistency is usually higher than 0.90
I1	Sufficient condition analysis	Identification of configurations leading to high or low critical thinking	Truth-table consistency cutoffs were set according to the fsQCA output, while configuration-level consistency and coverage were reported and interpreted cautiously
J1	fsQCA solution summary	Summary of solution coverage and consistency	Solution coverage and consistency were reported, and the configurations were interpreted cautiously

5 Empirical Analysis and Results

5.1 Common Method Bias, Descriptive Statistics, and Correlations

The results did not indicate a serious common method bias concern. Harman's single-factor test extracted six factors with eigenvalues above 1, and the first factor explained 24.86% of the variance, below the 40% threshold. The common method bias results are reported in Table 6. Full collinearity VIF values ranged from 1.18 to 1.57, below 3.3. The full collinearity VIF results are shown in Table 7. Descriptive statistics showed no substantial skewness or kurtosis problems. Correlations supported the dual-path logic: AIQ correlated positively with both SCO ($r = 0.410$) and SUCO ($r = 0.329$); SCO correlated positively with PRDP ($r = 0.222$), whereas SUCO correlated negatively with PRDP ($r = -0.380$). PRDP correlated positively with CT ($r = 0.245$), while SUCO correlated negatively with CT ($r = -0.252$). The near-zero correlation between AIQ and PRDP ($r = 0.001$) suggests possible offsetting effects between scaffolding and substitution. Descriptive statistics and correlations are reported in Table 8.

Table 6
Results of Common Method Bias Test

Code	Test method	Test indicator	Result	Judgment criterion	Conclusion
A1	Harman's single-factor test	Number of factors with eigenvalues greater than 1	6	Multifactor structure	Passed
B1	Harman's single-factor test	Variance explained by the first factor	24.86%	< 40%	No serious common method bias
C1	Full collinearity VIF test	Minimum VIF value	1.18	< 3.3	Passed
D1	Full collinearity VIF test	Maximum VIF value	1.57	< 3.3	Passed

Table 7
Full Collinearity VIF Test of Latent Variables

Code	Latent variable	Abbreviation	VIF
A1	Quality of conversational AI learning support	AIQ	1.42
B1	Scaffolded cognitive offloading	SCO	1.31
C1	Substitutive cognitive offloading	SUCO	1.57
D1	Post-response deep processing	PRDP	1.48
E1	AI evaluation literacy	AEL	1.18
F1	Critical thinking	CT	1.36

Table 8
Descriptive Statistics and Correlation Matrix (N = 426)

Code	Variable	Mean	SD	Skewness	Kurtosis	1	2	3	4	5	6
A1	AIQ	-0.00	1.00	0.053	-0.613	1.000					
B1	SCO	-0.00	1.00	0.006	-0.735	0.410	1.000				
C1	SUCO	-0.00	1.00	-0.011	-0.562	0.329	0.177	1.000			
D1	PRDP	-0.00	1.00	0.012	-0.628	0.001	0.222	-0.380	1.000		
E1	AEL	-0.00	1.00	0.076	-0.761	0.223	0.053	-0.014	0.064	1.000	
F1	CT	0.00	1.00	0.089	-0.690	-0.084	0.157	-0.252	0.245	0.003	1.000

5.2 Measurement Model

Most indicators reached acceptable reliability and convergent validity. AEL1 had a low loading (0.457), but it was retained because it represented an important dimension of AI evaluation literacy and its removal did not substantially improve construct-level reliability or convergent validity. Reliability and convergent validity results are reported in Table 9. From a content-validity perspective, retaining AEL1 preserved conceptual coverage while the AEL construct met accepted CA, CR, and AVE standards. Future research should further refine this item to improve the measurement precision of AI evaluation literacy. The self-developed or contextualized constructs also showed acceptable measurement quality: SCO (CA = 0.802, CR = 0.871, AVE = 0.627), SUCO (CA = 0.761, CR = 0.847, AVE = 0.581), and

PRDP (CA = 0.876, CR = 0.906, AVE = 0.617). A summary of the reliability and validity of the self-developed or contextually adapted constructs is presented in Table 10. HTMT values ranged from 0.051 to 0.481, below 0.85, supporting discriminant validity. The HTMT results are shown in Table 11. The HTMT value between SCO and SUCO was 0.221, indicating that the two forms of offloading are empirically distinct. The measurement model is shown in Fig. 4.

Table 9
Reliability and Convergent Validity of the Measurement Model (N = 426)

Code	Construct	Item	FL	IR	CA	CR	AVE
A1	AIQ	AIQ1	0.834	0.696	0.875	0.909	0.667
A2		AIQ2	0.815	0.664			
A3		AIQ3	0.774	0.599			
A4		AIQ4	0.798	0.637			
A5		AIQ5	0.861	0.741			
B1	SCO	SCO1	0.786	0.618	0.802	0.871	0.627
B2		SCO2	0.814	0.663			
B3		SCO3	0.752	0.566			
B4		SCO4	0.814	0.663			
C1	SUCO	SUCO1	0.755	0.570	0.761	0.847	0.581
C2		SUCO2	0.836	0.699			
C3		SUCO3	0.752	0.566			
C4		SUCO4	0.699	0.489			
D1	PRDP	PRDP1	0.758	0.575	0.876	0.906	0.617
D2		PRDP2	0.826	0.682			
D3		PRDP3	0.822	0.676			
D4		PRDP4	0.778	0.605			
D5		PRDP5	0.738	0.545			
D6		PRDP6	0.787	0.619			
E1	AEL	AEL1	0.457	0.209	0.856	0.842	0.525
E2		AEL2	0.779	0.607			
E3		AEL3	0.900	0.810			
E4		AEL4	0.724	0.524			
E5		AEL5	0.691	0.477			
F1	CT	CT1	0.805	0.648	0.890	0.916	0.646

Table 9. Reliability and Convergent Validity of the Measurement Model (N = 426)

Code	Construct	Item	FL	IR	CA	CR	AVE
F2		CT2	0.843	0.711			
F3		CT3	0.829	0.687			
F4		CT4	0.787	0.619			
F5		CT5	0.802	0.643			
F6		CT6	0.751	0.564			

Note: The self-developed constructs SCO, SUCO, and PRDP all reached acceptable levels in terms of CA, CR, and AVE, indicating that the self-developed scales had good internal consistency and convergent validity.

Table 10
Summary of Reliability and Validity of Self-Developed or Contextually Adapted Constructs

Code	Construct	Number of items	FL range	CA	CR	AVE	Conclusion
A1	SCO	4	0.752–0.814	0.802	0.871	0.627	Good reliability and convergent validity
B1	SUCO	4	0.699–0.836	0.761	0.847	0.581	Acceptable; SUCO4 was close to the threshold
C1	PRDP	6	0.738–0.826	0.876	0.906	0.617	Good reliability and convergent validity

Table 11
Discriminant Validity Test Results (HTMT)

Code	Construct	AEL	AIQ	CT	PRDP	SCO	SUCO
A1	AEL	—					
B1	AIQ	0.287	—				
C1	CT	0.055	0.102	—			
D1	PRDP	0.058	0.051	0.273	—		
E1	SCO	0.073	0.481	0.185	0.264	—	
F1	SUCO	0.055	0.389	0.298	0.458	0.221	—

Note: HTMT values below 0.85 or 0.90 generally indicate acceptable discriminant validity among constructs.

5.3 Structural Model and Mediation

The structural model supported H1 to H5 but not H8 and H9. The structural model path coefficients are reported in Table 12. AIQ was positively associated with SCO (β = 0.410, t = 10.524, p < .001, f^2 = 0.202) and SUCO (β = 0.329, t = 7.645, p < .001, f^2 = 0.122). SCO was positively associated with PRDP (β = 0.294, t = 6.663, p < .001, f^2 = 0.108), whereas SUCO was negatively associated with PRDP (β = -0.434, t = 11.372, p < .001, f^2 = 0.237). PRDP was positively associated with CT (β = 0.245, t = 5.301, p < .001, f^2 = 0.064). AEL was not significantly associated with PRDP (β = 0.044, p = 0.559), and the AEL × SUCO interaction was also nonsignificant (β = 0.029, p = 0.511). The standardized structural model results are shown in Fig. 5.

Bootstrapping results supported both indirect pathways. The indirect association of SCO with CT through PRDP was significant (β = 0.072, t = 4.212, p < .001, 95% CI [0.042, 0.110]), supporting H6. The indirect association of SUCO with CT through PRDP was also significant and negative (β = -0.106, t = 4.595, p < .001, 95% CI [-0.142, -0.057]), supporting H7. Bootstrap results for the specific indirect associations are reported in Table 13. The model explained 16.8% of SCO, 10.9% of SUCO, 23.4% of PRDP, and 6.0% of CT. The explanatory power of the endogenous constructs is summarized in Table 14. The modest explanatory power for CT indicates that additional learner-, task-, and instruction-level factors should be considered in future models.

Table 12
Structural Model Path Coefficients (N = 426)

Code	Path	β	t value	p value	f^2	Result
A1	AIQ → SCO	0.410	10.524	< .001	0.202	Significant positive association; H1 supported
B1	AIQ → SUCO	0.329	7.645	< .001	0.122	Significant positive association; H2 supported
C1	SCO → PRDP	0.294	6.663	< .001	0.108	Significant positive association; H3 supported
D1	SUCO → PRDP	-0.434	11.372	< .001	0.237	Significant negative association; H4 supported
E1	PRDP → CT	0.245	5.301	< .001	0.064	Significant positive association; H5 supported

Table 12. Structural Model Path Coefficients (N = 426)

Code	Path	β	t value	p value	f^2	Result
F1	AEL → PRDP	0.044	0.585	0.559	0.003	Nonsignificant direct association
G1	AEL × SUCO → PRDP	0.029	0.657	0.511	0.001	Nonsignificant interaction term; H8 not supported

Table 13
Bootstrap Results of Specific Indirect Effects

Code	Hypothesis	Specific indirect path	β	SE	t value	p value	95% bias-corrected bootstrap CI	Result
A1	H6	SCO → PRDP → CT	0.072	0.017	4.212	< .001	[0.042, 0.110]	Supported
B1	H7	SUCO → PRDP → CT	-0.106	0.023	4.595	< .001	[-0.142, -0.057]	Supported

Note: SCO = scaffolded cognitive offloading; SUCO = substitutive cognitive offloading; PRDP = post-response deep processing; CT = critical thinking. The confidence intervals are bias-corrected bootstrap confidence intervals.

Table 14
Explanatory Power of Endogenous Constructs

Code	Endogenous construct	Main predictors	R^2	Adjusted R^2	Interpretation
A1	SCO	AIQ	0.168	0.166	Positive explanatory power
B1	SUCO	AIQ	0.109	0.106	Positive explanatory power
C1	PRDP	AEL, SCO, SUCO, AEL × SUCO	0.234	0.226	Highest explanatory power
D1	CT	PRDP	0.060	0.058	Modest explanatory power

Note: R^2 and adjusted R^2 were extracted from the SmartPLS algorithm outputs. AIQ = quality of conversational AI learning support; SCO = scaffolded cognitive offloading; SUCO = substitutive cognitive offloading; PRDP = post-response deep processing; AEL = AI evaluation literacy; CT = critical thinking.

5.4 fsQCA and Importance–Performance Results

For fsQCA, continuous variables were calibrated using the 95th percentile as full membership, the 50th percentile as the crossover point, and the 5th percentile as full non-membership. The variable calibration anchors are reported in Table 15. No single condition reached the 0.90 consistency threshold for necessity, and the necessary condition test results are shown in Table 16. Truth-table analyses used a frequency threshold of 3, with consistency cutoffs of 0.808 for high CT and 0.802 for low CT. The high-CT configurations showed recurring presence of SCO and absence of SUCO. HC1 linked high CT with SCO, low SUCO, and PRDP even under relatively low AIQ; HC2 linked high CT with high AIQ, SCO, low SUCO, and high AEL. Low-CT configurations more often involved absence of SCO and presence of SUCO, sometimes combined with insufficient PRDP or low AEL. The typical configurational results are reported in Table 17 and illustrated in Fig. 6. Because several configuration-level and solution consistency values were below 0.80, these findings are interpreted as supplementary and exploratory configurational evidence rather than definitive causal pathways.

Table 15
Variable Calibration Anchors

Code	Variable	Abbreviation	Full membership point (0.95)	Crossover point (0.50)	Full non-membership point (0.05)
A1	AIQ	AIQs	0.773	0.501	0.208
B1	SCO	SCOs	0.787	0.501	0.213
C1	SUCO	SUCOs	0.787	0.501	0.187
D1	PRDP	PRDPs	0.796	0.501	0.204
E1	AEL	AELs	0.773	0.501	0.208
F1	CT	CTs	0.780	0.498	0.215

Table 16
Necessary Condition Test: Consistency and Coverage

Code	Condition	Consistency for high CT	Coverage for high CT	Consistency for low CT	Coverage for low CT
A1	AIQs	0.519	0.524	0.574	0.569
B1	~AIQs	0.573	0.578	0.520	0.515
C1	SCOs	0.596	0.596	0.504	0.495
D1	~SCOs	0.495	0.504	0.589	0.589
E1	SUCOs	0.494	0.493	0.627	0.614
F1	~SUCOs	0.613	0.626	0.482	0.484
G1	PRDPs	0.604	0.617	0.478	0.480
H1	~PRDPs	0.490	0.489	0.618	0.605
I1	AELs	0.538	0.553	0.538	0.544
J1	~AELs	0.556	0.551	0.557	0.542

Note: The calibrated variables were named AIQs, SCOs, SUCOs, PRDPs, AELs, and CTs. PRDPs correspond to post-response deep processing in this study.

Table 17
Typical Configurational Results for High and Low Critical Thinking

Code	Outcome	Config.	AIQ	SCO	SUCO	PRDP	AEL	Raw cov.	Unique cov.	Consist.
A1	High CT	HC1	⊗	●	⊗	●	—	0.173	0.106	0.788
B1	High CT	HC2	●	●	⊗	—	●	0.160	0.093	0.797
C1	Low CT	LC1	—	⊗	●	●	⊗	0.115	0.016	0.790
D1	Low CT	LC2	—	⊗	●	⊗	●	0.168	0.053	0.773
E1	Low CT	LC3	●	⊗	⊗	⊗	⊗	0.084	0.031	0.813
F1	Low CT	LC4	⊗	⊗	●	●	—	0.121	0.000	0.776
G1	Low CT	LC5	⊗	⊗	●	—	●	0.134	0.004	0.786

Note: ● indicates the presence or a high level of a condition; ⊗ indicates the absence or a low level of a condition; "—" indicates that the condition is not specified in the configuration. AIQ = quality of conversational AI learning support; SCO = scaffolded cognitive offloading; SUCO = substitutive cognitive offloading; PRDP = post-response deep processing; AEL = AI evaluation literacy; CT = critical thinking.

The fsQCA solution coverage and consistency were 0.266 and 0.775 for high CT and 0.281 and 0.747 for low CT, indicating that the configurational evidence should be interpreted cautiously. The fsQCA solution parameters are summarized in Table 18. IPMA further showed that PRDP had the highest importance for CT (total effect = 0.245, performance = 49.897), SUCO had a negative total effect (-0.106), and SCO had a positive but modest total effect (0.072). The IPMA results are reported in Table 19. Within the current model, this pattern suggests that support quality and evaluation literacy alone were not strongly related to critical thinking; learners' continued verification, comparison, questioning, reflection, and revision showed a more direct association with critical thinking. The importance–performance map for critical thinking is shown in Fig. 7.

Table 18
Summary of fsQCA Solutions and Analytical Parameters

Code	Outcome variable	Frequency threshold	Truth-table consistency cutoff	Number of configurations	Solution coverage	Solution consistency
A1	High CT	3	0.808	2	0.266	0.775
B1	Low CT	3	0.802	5	0.281	0.747

Table 19
Importance–Performance Matrix Analysis Results: Target Variable CT

Code	Variable	Importance, total effect	Performance
A1	Post-response deep processing	0.245	49.897
B1	Substitutive cognitive offloading	-0.106	49.820
C1	Scaffolded cognitive offloading	0.072	50.138
D1	AI evaluation literacy	0.011	49.838
E1	Quality of conversational AI learning support	-0.006	50.003

6 Discussion

6.1 Major Findings

The results provide preliminary support for the dual-path assumption. Conversational AI learning support was associated with both scaffolded and substitutive offloading. In the scaffolding pathway, AIQ was associated with SCO, which was positively related to PRDP and, indirectly, CT. In the substitution pathway, AIQ was associated with SUCO, and SUCO was negatively related to PRDP, which in turn was associated with lower CT. These opposing mechanisms may explain why the overall association between support quality and critical thinking was weak. PRDP emerged as the proximal explanatory process, suggesting that learners' continued active monitoring is central to whether conversational AI use is associated with more favorable critical thinking outcomes.

The nonsignificant moderation of AEL suggests that evaluative knowledge alone may not automatically become verification behavior. Learners may understand that generated responses contain errors or biases but still rely on them for convenience, efficiency, or perceived authority. The fsQCA results provide a complementary view: AEL may matter more as part of specific configurations than as a simple linear moderator. High CT tended to appear when learners showed scaffolded offloading and low substitutive offloading, whereas low CT was more often associated with insufficient scaffolding, increased substitution, and weak post-response processing. However, given the modest R^2 for CT and the exploratory consistency of fsQCA solutions, these findings should be interpreted as evidence for a specific cognitive mechanism rather than as a comprehensive prediction model.

6.2 Contributions and Implications

The main theoretical contribution of this study is the distinction between two forms of cognitive offloading. Rather than treating AI use as uniformly supportive or harmful, the findings show that scaffolded and substitutive offloading are associated with different patterns of post-response processing. The study also highlights post-response deep processing as a proximal process linking AI-generated responses with critical thinking. In addition, the nonsignificant

moderation result suggests that AI evaluation literacy may need to be examined together with behavioral and contextual conditions rather than only as a linear moderator.

For teaching practice, the findings point to the value of tasks that ask students to justify, compare, and revise generated responses rather than accept them as final answers. Reflective prompts, verification tasks, comparison activities, and case-based discussions may help students maintain cognitive vigilance. For system design, the results suggest the need to move beyond answer generation by adding prompts that challenge learners, mark uncertainty, request evidence checking, and support follow-up reasoning. AI-supported learning systems may therefore be evaluated not only by efficiency or perceived usefulness, but also by the extent to which they encourage learners to sustain cognitive responsibility after responses are generated.

7 Limitations and Future Research Directions

Several limitations should be acknowledged. First, the sample was mainly drawn from one university, so future studies should test the model across educational stages, disciplines, and institutions. Second, the cross-sectional self-report design limits causal interpretation; experimental designs, behavioral indicators, or longitudinal tracking would support stronger causal interpretations. Third, future work could compare text-based and multimodal tools and examine how interaction turns, feedback types, and prompt strategies shape cognitive offloading and critical thinking (Hutler et al., 2025). Finally, the measurement of AI evaluation literacy, especially AEL1, should be refined.

8 Conclusion

This study suggests that the relationship between conversational AI learning support and critical thinking depends not only on response quality, but also on how learners process AI-generated responses after receiving them. This argument is consistent with recent research indicating that generative AI may reshape cognitive offloading and learning processes (Yin et al., 2025; Seung & Basham, 2026). The PLS-SEM results showed opposite scaffolding and substitution pathways, and the fsQCA results provided supplementary evidence that high critical thinking was more likely when scaffolded offloading was present and substitutive offloading was low. The findings should be interpreted cautiously because CT explained variance was modest and AEL moderation was nonsignificant. Overall, the findings suggest that conversational AI may have greater educational value when learners are encouraged to verify, question, compare, reflect on, and revise AI-generated responses rather than simply accept them.

Declarations

Data Availability

The individual-level questionnaire data generated and/or analysed during the current study are not publicly available because the data were collected under an approved ethics protocol requiring confidential handling of participant responses. Although no direct identifiers were collected, the dataset includes demographic characteristics and AI-use background information that may create a residual risk of participant re-identification when combined. Aggregate results supporting the findings are reported in the article, and the full measurement items and scoring scheme are provided in Supplementary Material 1. De-identified individual-level data are available from the corresponding author, Zhentai Zhang (2601888596@qq.com), upon reasonable request for non-commercial research purposes, subject to consistency with the approved ethics protocol and a data-use agreement prohibiting participant re-identification, redistribution, or onward sharing.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding Statement

This research was supported by the Department of Education of Liaoning Province. The funder had no role in the study design, data collection, data analysis, interpretation of the results, or preparation of the manuscript.

Author Contributions

Yiding Bu contributed to the conceptualization and design of the study, supervised the research process, critically revised the manuscript for important intellectual content, and approved the final version for publication. **Zhentai Zhang** contributed to the study design, data collection, data analysis, interpretation of the results, and drafting of the manuscript. **Yuxin Li** contributed to data collection, questionnaire preparation, literature organization, and manuscript revision. **Shuyi Wang** contributed to the interpretation of the findings, critical revision of the manuscript, and methodological refinement. All authors approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

References

1. Akolekar, H. et al. The role of generative AI tools in shaping mechanical engineering education from an undergraduate perspective. *Sci. Rep.* **15** (1), 9214. https://doi.org/10.1038/s41598-025-93871-z (2025).
2. Hu, N. et al. Status and perceptions of ChatGPT utilization among medical students: a survey-based study. *BMC Med. Educ.* **25** (1), 831. https://doi.org/10.1186/s12909-025-07438-7 (2025).
3. Qian, Y. Pedagogical applications of generative AI in higher education: A systematic review of the field. *TechTrends*, 1–16. (2025). https://doi.org/10.1007/s11528-025-01100-1
4. Nasr, N. R. et al. Exploring the Impact of Generative AI ChatGPT on Critical Thinking in Higher Education: Passive AI-Directed Use or Human–AI Supported Collaboration? *Educ. Sci.* **15** (9), 1198. https://doi.org/10.3390/educsci15091198 (2025).
5. Gerlich, M. AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies* **15** (1), 6. https://doi.org/10.3390/soc15010006 (2025).
6. Hu, Y., Curdt-Christiansen, X. L. & Wang, J. Smart tools, smarter minds? Learner-AI interaction and AI assistance on critical thinking in EAP contexts. *J. Engl. Acad. Purp.* **78**, 101586. https://doi.org/10.1016/j.jeap.2025.101586 (2025).
7. Arthi, M. P. & Gandhimathi, S. N. S. Research trends and network approach of critical thinking skills in English Language Teaching–A bibliometric analysis implementing R studio. *Heliyon* **11** (2). https://doi.org/10.1016/j.heliyon.2025.e42080 (2025).
8. Shafer, D. A critical thinking thematic framework and observation tool for improved theory and developing secondary teachers' instructional practice: Proof of concept. *Think. Skills Creativity*. **56**, 101787. https://doi.org/10.1016/j.tsc.2025.101787 (2025).
9. Nguyen, H. Role design considerations of conversational agents to facilitate discussion and systems thinking. *Comput. Educ.* **192**, 104661. https://doi.org/10.1016/j.compedu.2022.104661 (2023).
10. Körner, A., Tolzin, A., Janson, A., Leimeister, J. M. & Rummer, R. Common ground improves learning with conversational agents. *Behav. Inform. Technol.* **45** (5), 932–948. https://doi.org/10.1080/0144929X.2025.2541222 (2026).

11. Ortega-Ochoa, E., Arguedas, M. & Daradoumis, T. Empathic pedagogical conversational agents: a systematic literature review. *Br. J. Edu. Technol.* **55** (3), 886–909. https://doi.org/10.1111/bjet.13413 (2024).
12. Hou, C. et al. The Effects of Critical Thinking Intervention on Reliance Behaviors, Problem-solving Quality, and Creativity during Human-Generative AI Collaborative Learning. *Comput. Educ.* 105576. https://doi.org/10.1016/j.compedu.2026.105576 (2026).
13. Guo, Y. & Ye, Q. Meta-cognitive insights into cognitive offloading: mechanisms, interventions, and educational implications. *Humanit. Social Sci. Commun.* https://doi.org/10.1057/s41599-026-06621-5 (2026).
14. Vendrell, M. & Johnston, S. K. Scaffolding critical thinking with generative AI: Design principles for integrating large language models in higher education. *Computers Education: Artif. Intell.* 100572. https://doi.org/10.1016/j.caeai.2026.100572 (2026).
15. Anders, A. D. & Speltz, E. D. Developing generative AI literacies through self-regulated learning: A human-centered approach. *Computers Education: Artif. Intell.* 100482. https://doi.org/10.1016/j.caeai.2025.100482 (2025).
16. Yavich, R. Will the use of AI undermine students independent thinking? *Educ. Sci.* **15** (6), 669. https://doi.org/10.3390/educsci15060669 (2025).
17. Peng, N., Logie, R. H. & Della Sala, S. Effect of levels-of-processing on rates of forgetting. *Mem. Cognit.* **53** (2), 692–709. https://doi.org/10.3758/s13421-024-01581-4 (2025).
18. Luchini, S. A., Wang, S., Kenett, Y. N. & Beaty, R. E. Mapping the memory structure of high-knowledge students: A longitudinal semantic network analysis. *J. Intell.* **12** (6), 56. https://doi.org/10.3390/jintelligence12060056 (2024).
19. Loyens, S. M., Van Meerten, J. E., Schaap, L. & Wijnia, L. Situating higher-order, critical, and critical-analytic thinking in problem-and project-based learning environments: A systematic review. *Educational Psychol. Rev.* **35** (2), 39. https://doi.org/10.1007/s10648-023-09757-x (2023).
20. Tan, A. J., Davies, J. L., Nicolson, R. I. & Karaminis, T. Learning critical thinking skills online: can precision teaching help? *Education Tech. Research Dev.* **71** (3), 1275–1296. https://doi.org/10.1007/s11423-023-10227-y (2023).
21. Balart, T., Díaz, B. & Shryock, K. A Systematic Literature Review on the Pedagogical Implications and Impact of GenAI on Students' Critical Thinking. *Algorithms* **19** (3), 179. https://doi.org/10.3390/a19030179 (2026).
22. Zhou, X., Teng, D. & Al-Samarraie, H. The mediating role of generative AI self-regulation on students' critical thinking and problem-solving. *Educ. Sci.* **14** (12), 1302. https://doi.org/10.3390/educsci14121302 (2024).
23. Tang, K. S. & Putra, G. B. S. Generative AI as a dialogic partner: Enhancing multiple perspectives, reasoning, and argumentation in science education with customized chatbots. *J. Sci. Edu. Technol.* 1–13. https://doi.org/10.1007/s10956-025-10240-1 (2025).
24. Dong, Y., Xu, W., Huang, J. & Yann, K. Validating and refining a multi-dimensional scale for measuring AI literacy in education using the Rasch Model. *Humanit. Social Sci. Commun.* **12** (1), 1–13. https://doi.org/10.1057/s41599-025-05670-6 (2025).
25. Kim, W. & Ryoo, Y. The literacy paradox: How AI literacy amplifies biases in evaluating AI-generated news articles. *Int. J. Inf. Manag.* **86**, 102992. https://doi.org/10.1016/j.ijinfomgt.2025.102992 (2026).
26. Shen, Q. Same AI, Different Pathways: Unpacking Mechanisms of AI-Mediated Learning across Discipline-Institution Contexts. *Computers Education: Artif. Intell.* 100601. https://doi.org/10.1016/j.caeai.2026.100601 (2026).
27. Badali, M., Shahalizadeh, M., Alimirzaie, M., Noroozi, O. & Kazem Banihashem, S. Bridging AI chatbot use and critical thinking: the mediating role of AI literacy. *Interact. Learn. Environ.* 1–13. https://doi.org/10.1080/10494820.2025.2605487 (2026).
28. Yan, W. et al. Generative AI literacy: Scale development and its influence on privacy protection behaviors and information verification behaviors. *Telecomm. Policy*. 103117. https://doi.org/10.1016/j.telpol.2025.103117

(2026).

29. Zhai, C., Wibowo, S. & Li, L. D. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learn. Environ.* **11** (1), 28. https://doi.org/10.1186/s40561-024-00316-7 (2024).
30. Li, C., Cui, H. & Hagedorn, L. S. The cognitive impact of ChatGPT in higher education: A systematic review of critical and creative thinking outcomes. *Computers Education: Artif. Intell.* 100571. https://doi.org/10.1016/j.caeai.2026.100571 (2026).
31. Gerlich, M. From offloading to engagement: an experimental study on structured prompting and critical reasoning with generative AI. *Data* **10** (11), 172. https://doi.org/10.3390/data10110172 (2025).
32. Iqbal, J., Hashmi, Z. F., Asghar, M. Z. & Abid, M. N. Generative AI tool use enhances academic achievement in sustainable education through shared metacognition and cognitive offloading among preservice teachers. *Sci. Rep.* **15** (1), 16610. https://doi.org/10.1038/s41598-025-01676-x (2025).
33. Lin, Y. R., Jiao, X. & Hung, C. Y. Enhancing scientific argumentation with visual scaffolding in generative AI-supported environments. *Int. J. Sci. Educ.* 1–26. https://doi.org/10.1080/09500693.2026.2630293 (2026).
34. Wang, J. Cognitive offloading through digital tools and its relationship with critical thinking, task persistence, and learning depth. *Front. Psychol.* **17**, 1781101. https://doi.org/10.3389/fpsyg.2026.1781101 (2026).
35. Yang, Y., Zhang, Y., Sun, D., He, W. & Wei, Y. Navigating the landscape of AI literacy education: insights from a decade of research (2014–2024). *Humanit. Social Sci. Commun.* **12** (1), 1–12. https://doi.org/10.1057/s41599-025-04583-8 (2025).
36. Du, X., Du, M., Zhou, Z. & Bai, Y. Facilitator or hindrance? The impact of AI on university students' higher-order thinking skills in complex problem solving. *Int. J. Educational Technol. High. Educ.* **22** (1), 39. https://doi.org/10.1186/s41239-025-00534-0 (2025).
37. Meng, N., Deli, M., Abdul Rauf, U. A. & M., & Educational Technology and AI: Bridging Cognitive Load and Learner Engagement for Effective Learning. *Sage Open.* **15** (4), 21582440251395930. https://doi.org/10.1177/21582440251395930 (2025).
38. Klingbeil, A., Grützner, C. & Schreck, P. Trust and reliance on AI—An experimental study on the extent and costs of overreliance on AI. *Comput. Hum. Behav.* **160**, 108352. https://doi.org/10.1016/j.chb.2024.108352 (2024).
39. Visser, R., Peters, T. M., Scharlau, I. & Hammer, B. Trust, distrust, and appropriate reliance in (X) AI: A conceptual clarification of user trust and survey of its empirical evaluation. *Cogn. Syst. Res.* **91**, 101357. https://doi.org/10.1016/j.cogsys.2025.101357 (2025).
40. Liu, J. et al. Adaptive vs. planned metacognitive scaffolding for computational thinking: Evidence from generative AI-supported programming in elementary education. *Comput. Educ.* 105473. https://doi.org/10.1016/j.compedu.2025.105473 (2025).
41. Tian, J. & Zhang, R. Outsourcing thinking to AI? Focused immersion, AI dependency, and the double-edged impact on critical thinking. *Humanit. Social Sci. Commun.* https://doi.org/10.1057/s41599-026-07153-8 (2026).
42. Stadler, M., Bannert, M. & Sailer, M. Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Comput. Hum. Behav.* **160**, 108386. https://doi.org/10.1016/j.chb.2024.108386 (2024).
43. Lee, H. Y., Chen, P. H., Wang, W. S., Huang, Y. M. & Wu, T. T. Empowering ChatGPT with guidance mechanism in blended learning: Effect of self-regulated learning, higher-order thinking skills, and knowledge construction. *Int. J. Educational Technol. High. Educ.* **21** (1), 16. https://doi.org/10.1186/s41239-024-00447-4 (2024).
44. Wang, W. S., Lin, C. J., Lee, H. Y., Huang, Y. M. & Wu, T. T. Enhancing self-regulated learning and higher-order thinking skills in virtual reality: the impact of ChatGPT-integrated feedback aids. *Educ. Inform. Technol.* **30** (14), 19419–19445. https://doi.org/10.1007/s10639-025-13557-x (2025).

45. Melanou, C. & Beege, M. Scaffolding generative AI as a tutor: A quasi-experimental study of learning outcomes and motivational, cognitive and metacognitive processes. *Educ. Sci.* **16** (4), 651. https://doi.org/10.3390/educsci16040651 (2026).
46. Li, W., Feng, Q., Zhu, X., Yu, Q. & Wang, Q. Effect of summarizing scaffolding and textual cues on learning performance, mental model, and cognitive load in a virtual reality environment: An experimental study. *Comput. Educ.* **200**, 104793. https://doi.org/10.1016/j.compedu.2023.104793 (2023).
47. Tian, J. & Zhang, R. Learners' AI dependence and critical thinking: The psychological mechanism of fatigue and the social buffering role of AI literacy. *Acta. Psychol.* **260**, 105725. https://doi.org/10.1016/j.actpsy.2025.105725 (2025).
48. Bećirović, S., Polz, E. & Tinkel, I. Exploring students' AI literacy and its effects on their AI output quality, self-efficacy, and academic performance. *Smart Learn. Environ.* **12** (1), 29. https://doi.org/10.1186/s40561-025-00384-3 (2025).
49. Clark, K. R. Verifying Outputs in Scopus AI to Promote Critical AI Literacy in Education and Research. *Radiol. Technol.* **97** (5), 303–309 (2026).
50. Zhu, Y., Liu, Q. & Zhao, L. Exploring the impact of generative artificial intelligence on students' learning outcomes: A meta-analysis. *Educ. Inform. Technol.* **30** (11), 16211–16239. https://doi.org/10.1007/s10639-025-13420-z (2025).
51. Zhong, B. & Liu, X. Evaluating AI literacy of secondary students: Framework and scale development. *Comput. Educ.* **227**, 105230. https://doi.org/10.1016/j.compedu.2024.105230 (2025).
52. Wang, Y. The effect of generative artificial intelligence literacy on academic achievement: the mediating role of academic self-efficacy and the moderating role of critical thinking. *Front. Psychol.* **17**, 1761562. https://doi.org/10.3389/fpsyg.2026.1761562 (2026).
53. Wang, F., Cheung, A. C., Chai, C. S. & Liu, J. Development and validation of the perceived interactivity of learner-AI interaction scale. *Educ. Inform. Technol.* **30** (4), 4607–4638. https://doi.org/10.1007/s10639-024-12963-x (2025).
54. Markus, A., Carolus, A. & Wienrich, C. Objective measurement of AI literacy: development and validation of the AI competency objective scale (AICOS). *Computers and Education: Artificial Intelligence*, 100485. (2025). https://doi.org/10.1016/j.caeai.2025.100485
55. Wang, Y. Y. & Chuang, Y. W. Artificial intelligence self-efficacy: Scale development and validation. *Educ. Inform. Technol.* **29** (4), 4785–4808. https://doi.org/10.1007/s10639-023-12015-w (2024).
56. Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F. & Chu, S. K. W. Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *Br. J. Edu. Technol.* **55** (3), 1082–1104. https://doi.org/10.1111/bjet.13411 (2024).
57. Lintner, T. A systematic review of AI literacy scales. *npj Sci. Learn.* **9** (1), 50. https://doi.org/10.1038/s41539-024-00264-4 (2024).
58. Yuan, C. W., Tsai, H. Y. S. & Chen, Y. T. Charting competence: A holistic scale for measuring proficiency in artificial intelligence literacy. *J. Educational Comput. Res.* **62** (7), 1455–1484. https://doi.org/10.1177/07356331241261206 (2024).
59. Rodríguez-Rojas, M. P., Ramírez-Segura, A. M., Valenzuela-Mora, P., Dampuré, J. & Riveros Munévar, F. Critical thinking evaluation scale: Design and validation in a Colombian population. *Sage Open.* **14** (4), 21582440241297418. https://doi.org/10.1177/21582440241297418 (2024).
60. Li, X. & Liu, J. Validating a critical thinking ability questionnaire for EFL learners. *Think. Skills Creativity*. **51**, 101442. https://doi.org/10.1016/j.tsc.2023.101442 (2024).

61. Richter, N. F. & Tudoran, A. A. Elevating theoretical insight and predictive accuracy in business research: Combining PLS-SEM and selected machine learning algorithms. *J. Bus. Res.* **173**, 114453. https://doi.org/10.1016/j.jbusres.2023.114453 (2024).
62. Grunwald, G., Kara, A. & Spillan, J. E. Fuzzy Set Qualitative Comparative Analysis (fsQCA) to Examine Causal Complexity in Business Students' Sustainability Behavioural Intentions. *High. Educ. Q.* **79** (4), e70076. https://doi.org/10.1111/hequ.70076 (2025).
63. Hutler, B. et al. Metacognitive AI literacy: findings from an interactive AI fair. *AI Soc.* 1–16. https://doi.org/10.1007/s00146-025-02803-y (2025).
64. Yin, J., Xu, H., Pan, Y. & Hu, Y. Effects of different AI-driven Chatbot feedback on learning outcomes and brain activity. *npj Sci. Learn.* **10** (1), 17. https://doi.org/10.1038/s41539-025-00311-8 (2025).
65. Seung, Y. & Basham, J. D. Cognitive Offloading in the Age of Generative AI: What Does It Mean for Students With Learning Disabilities? *Learn. Disabil. Q.* 07319487261439132. https://doi.org/10.1177/07319487261439132 (2026).

Figures

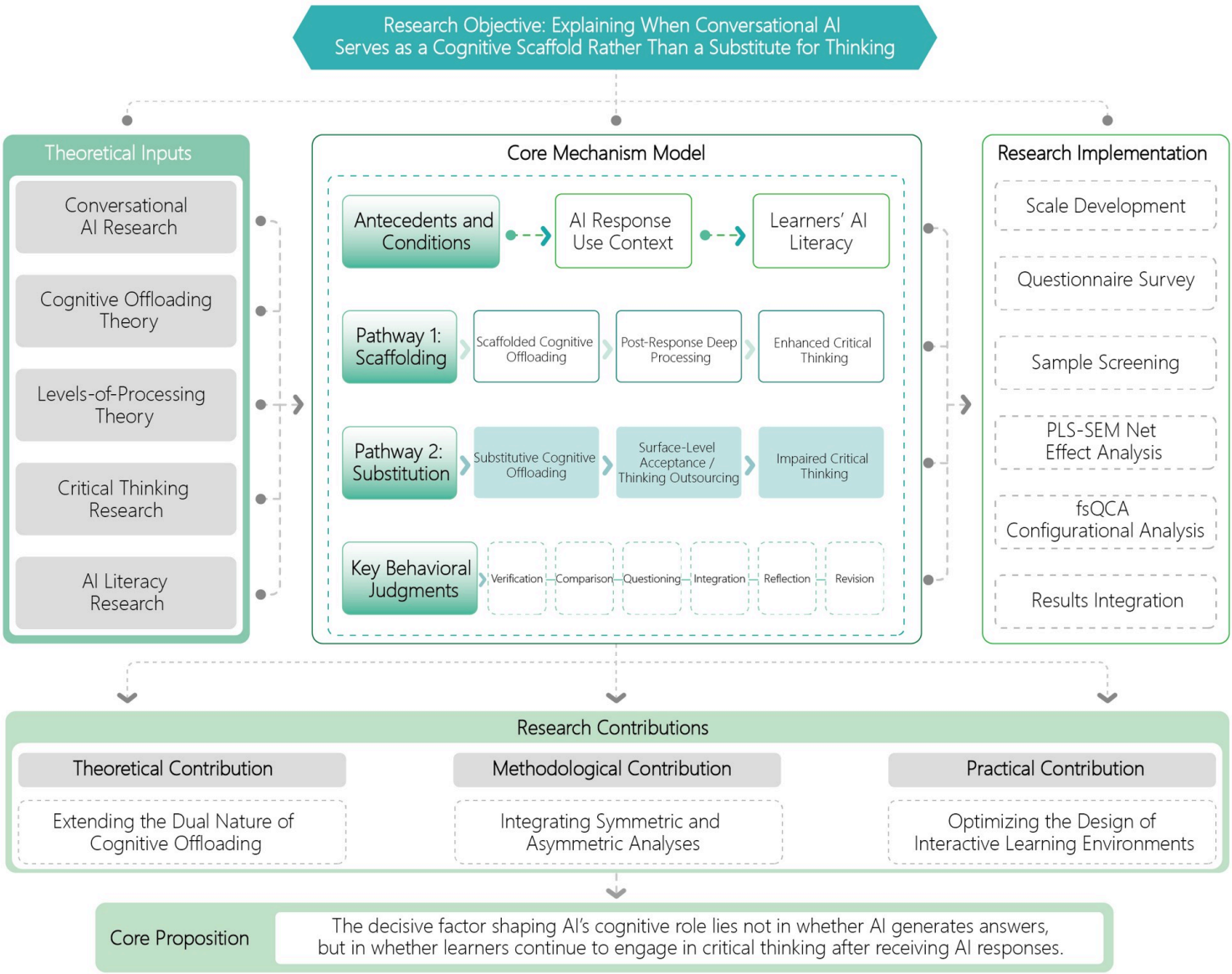


Figure 1

Research Process Framework

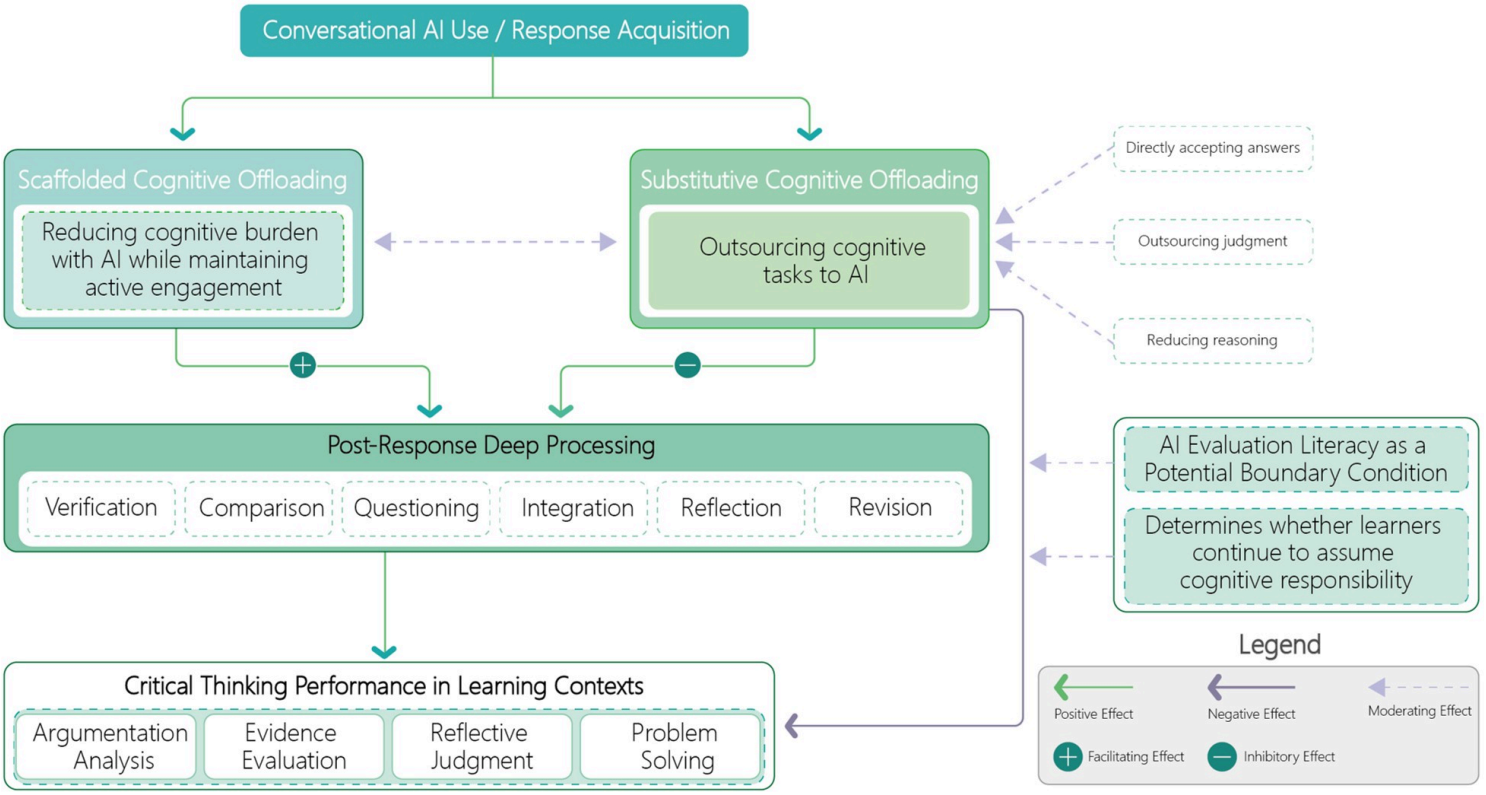


Figure 2

Structure of the Theoretical Model

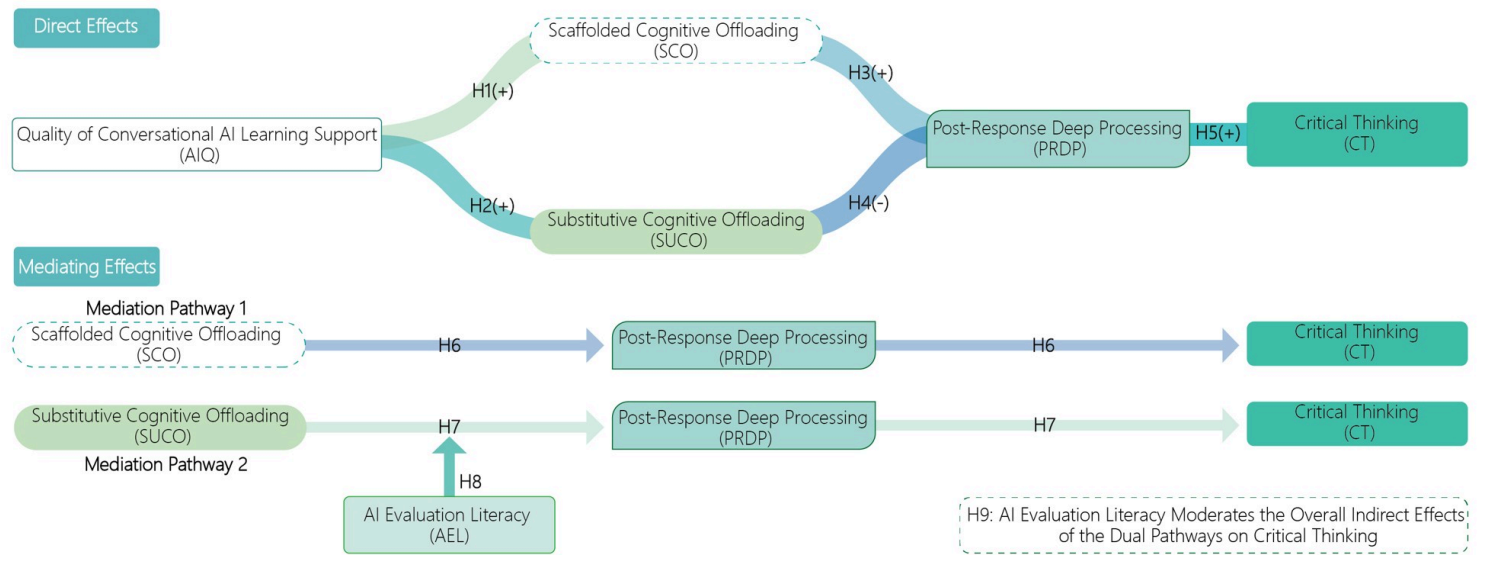


Figure 3

Structural Path Model

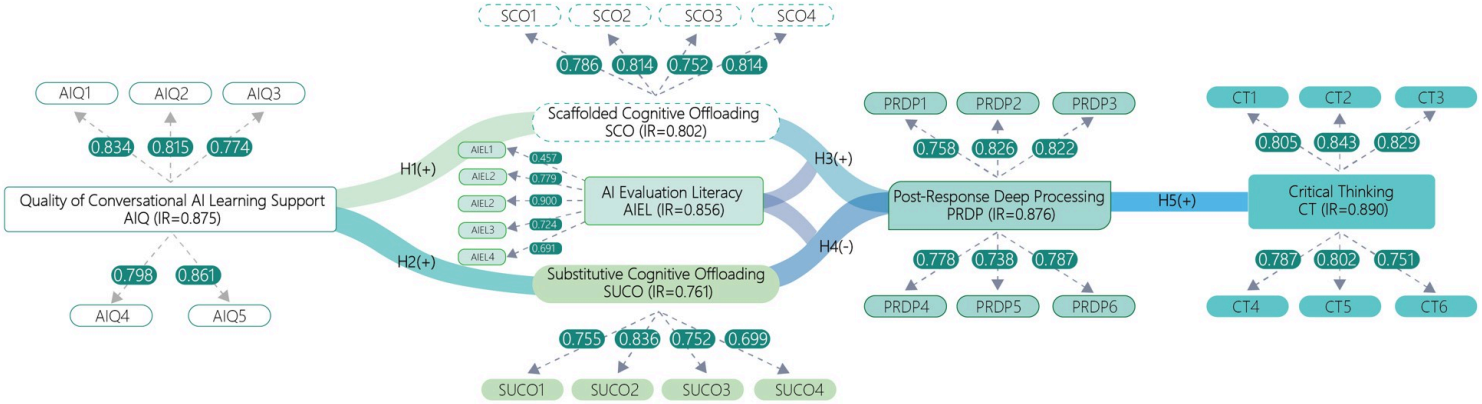


Figure 4

Measurement Model with Factor Loadings and Cronbach's Alpha Coefficients

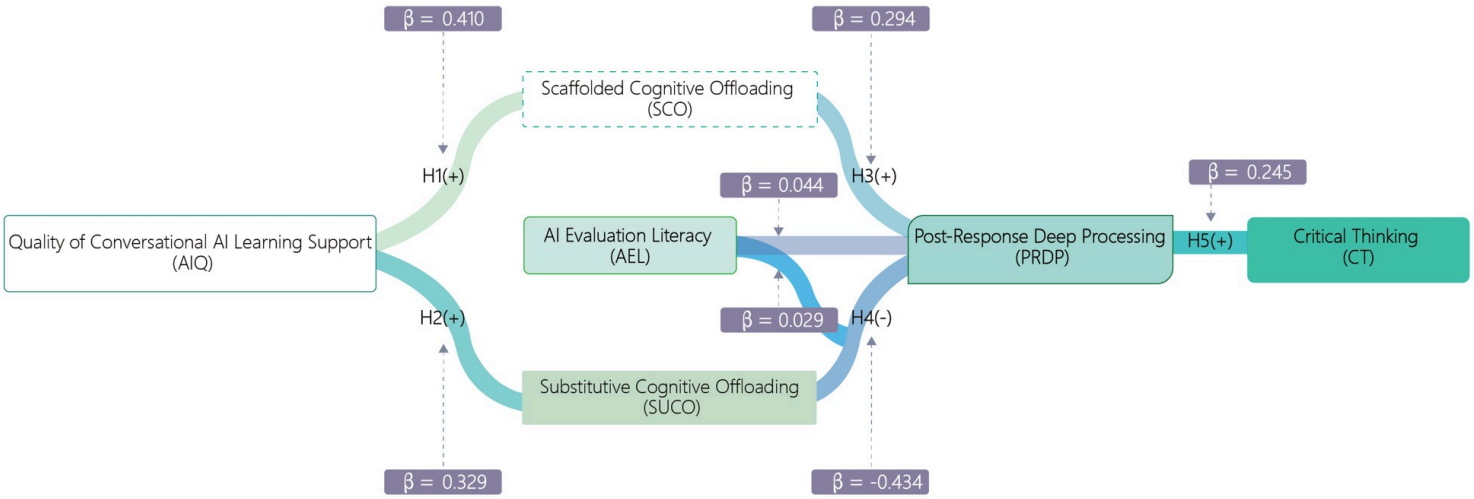


Figure 5

Standardized Structural Model Results

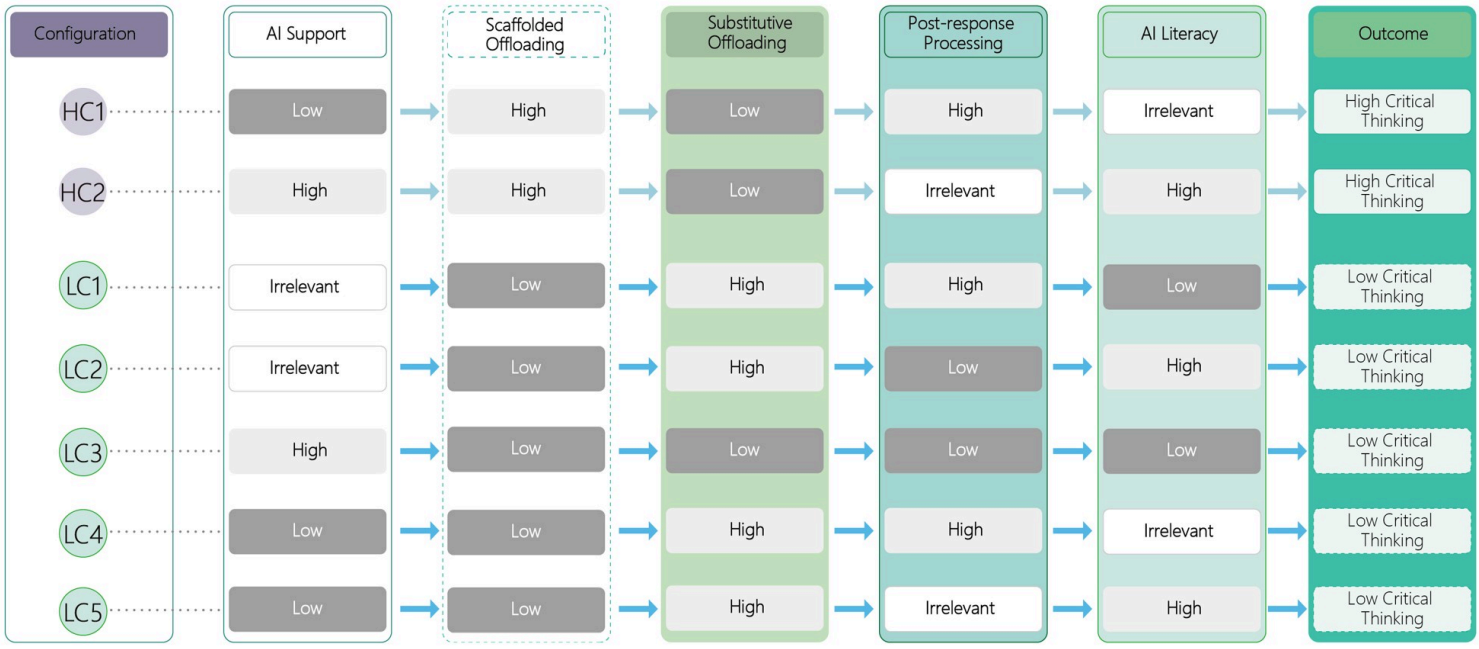


Figure 6

Diagram of Typical fsQCA Configurations

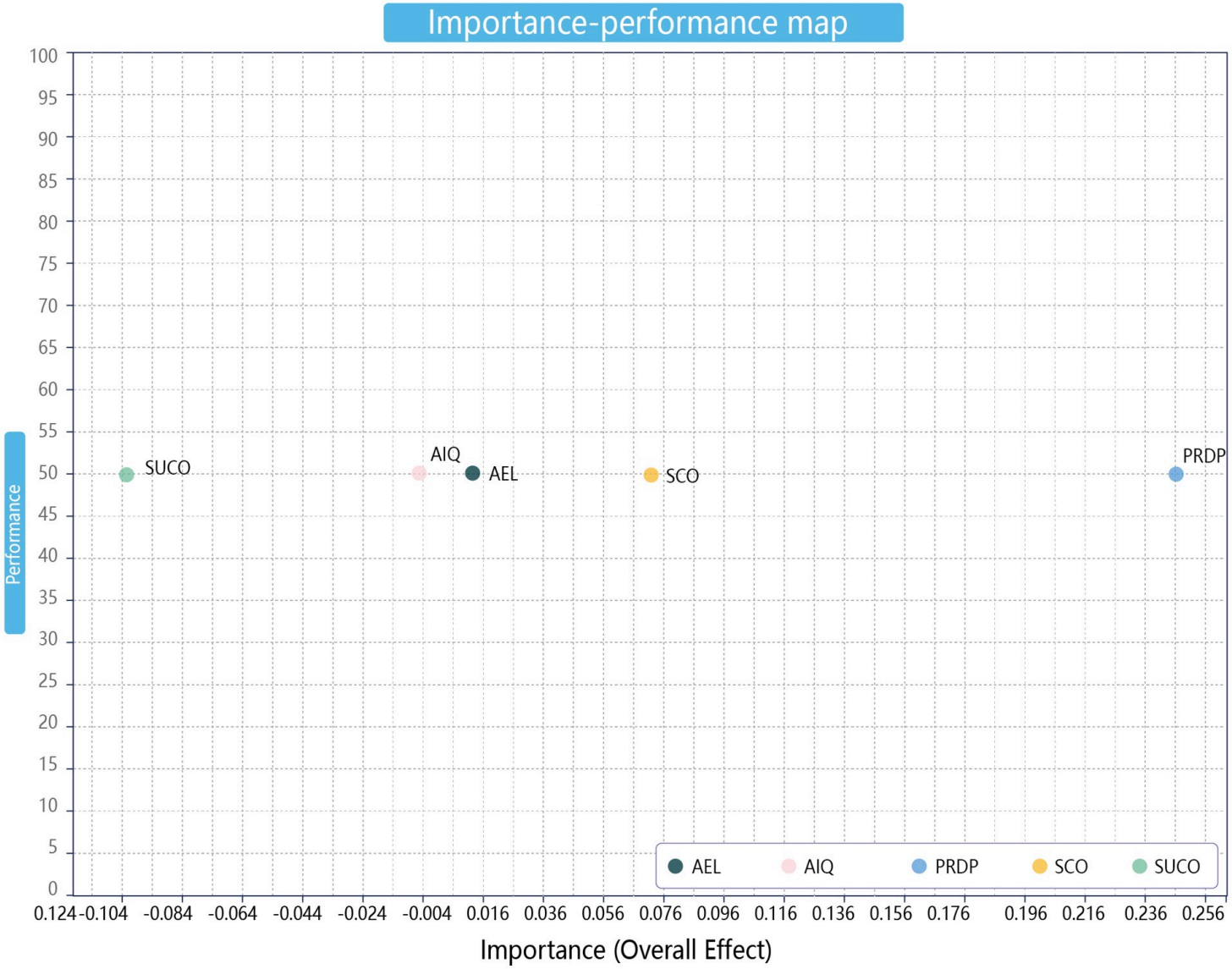


Figure 7

Importance–Performance Map for Critical Thinking at the Construct Level

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryMaterial1.docx