

From Assistance to Dependence: The Influence of AI-Generated Reading Texts on Critical Thinking Development among CPGE and EMSI Students in Morocco

*Adil DAHHAN¹ and Bani KOUMACHI²

¹Doctoral Student, FLLA, Ibn Tofail University, Kenitra, Morocco

²Associate Professor, FLLA, Ibn Tofail University, Kenitra, Morocco

*Corresponding Author

DOI: <https://doi.org/10.47772/IJRISS.2026.1026EDU0597>

Received: 17 September 2026; Accepted: 22 September 2026; Published: 29 September 2026

ABSTRACT

This study examines whether AI-generated reading texts can scaffold critical-thinking development in two academically demanding EFL contexts in Tangier, Morocco: the Classes Préparatoires aux Grandes Écoles (CPGE) and the École Marocaine des Sciences d'Ingénieurs (EMSI). Across an eight-week comparative sequence, performance under AI-generated and traditional authentic-text conditions was tracked at Weeks 1, 4, and 8. Scores rose from 54 to 82 in the AI-text condition and from 55 to 69 in the traditional-text condition. The AI condition therefore showed a 28-point gain, compared with a 14-point gain in the traditional condition; the descriptive difference-in-differences was 14 points. The same direction of difference appeared in both CPGE and EMSI. The findings suggest that AI-generated texts can function as useful cognitive scaffolds when paired with teacher mediation, higher-order reading tasks, and continued exposure to authentic complexity.

Keywords: AI-generated texts; critical thinking; EFL reading; cognitive load; scaffolding; CPGE; EMSI; Morocco

INTRODUCTION

Generative artificial intelligence has rapidly entered language education, giving teachers the capacity to produce, adapt, simplify and reorganise reading materials in ways that were previously time-consuming. This affordance is especially attractive in academically demanding EFL settings where students may understand the literal content of a text but still struggle to analyse arguments, infer unstated meanings and evaluate claims. Yet the pedagogical value of simplification is not self-evident. If AI-generated texts reduce unnecessary linguistic difficulty, they may release cognitive resources for higher-order thinking. If, however, they remove too much complexity or perform too much interpretive work for learners, they may reduce the productive cognitive effort on which critical thinking depends.

This tension is particularly relevant to two post-secondary settings in Tangier, Morocco. The first is the Classes Préparatoires aux Grandes Écoles (CPGE), a French-model preparatory system that prepares high-achieving students for competitive entry to engineering and management schools. The second is the École Marocaine des Sciences d'Ingénieurs (EMSI), a Moroccan engineering school with a campus in Tangier. The first author teaches English in both settings, and the research problem emerged from sustained classroom observation of students who generally demonstrated adequate comprehension but weaker performance when asked to analyse, infer and evaluate. The two settings therefore provide a contextually rich basis for examining whether text design can influence the development of critical reading.

The study is original in four related respects. First, it focuses on AI-generated reading texts as instructional materials rather than on students' direct use of ChatGPT as an answer-producing tool. Second, it compares AI-generated texts with authentic argumentative texts rather than comparing AI use with no technology. Third, it

follows development across an eight-week period instead of relying only on one-time perceptions. Fourth, it examines the same pedagogical question across CPGE and engineering-school contexts in Morocco, where empirical research on generative AI and critical reading remains limited. Accordingly, the study asks not whether AI is inherently beneficial or harmful, but under what instructional conditions it can function as scaffolding without encouraging dependence.

FOCUSED LITERATURE REVIEW

Critical thinking in EFL reading

Critical thinking involves purposeful, reasoned judgment rather than the passive acceptance of information. Ennis (1985) describes it as reasonable reflective thinking directed toward deciding what to believe or do, while Facione (1990) identifies interpretation, analysis, evaluation, inference, explanation and self-regulation as central skills. These processes are directly relevant to advanced reading. Readers must not only understand propositions but also establish relationships among claims, identify assumptions, draw warranted inferences and evaluate the adequacy of evidence. In second-language contexts, such processes operate alongside lexical, syntactic and discourse processing, which can place additional demands on working memory (Grabe, 2009; Kintsch, 1998).

Cognitive load and pedagogical scaffolding

Cognitive load theory provides one explanation for why text accessibility may affect critical performance. Working memory is limited, and learning can be impaired when excessive processing is devoted to demands that are not central to the intended learning goal (Sweller, 1988). From this perspective, clearer organisation and controlled linguistic difficulty can function as scaffolding by reducing extraneous load while preserving the need to analyse and evaluate. The crucial distinction is therefore between reducing unnecessary difficulty and removing productive intellectual challenge. Authentic texts remain valuable because they preserve rhetorical density, ambiguity and naturally occurring complexity; AI-generated texts may be useful when they help learners reach those higher-order demands rather than bypass them.

Generative AI, critical thinking, and the assistance-dependence tension

Recent research presents a conditional rather than uniformly positive account of generative AI. Essien et al. (2024) found that AI text generators could support aspects of critical thinking, although gains were stronger at lower cognitive levels and concerns remained about accuracy and overreliance. In an eight-week EFL intervention, Liu and Wang (2024) reported stronger critical-thinking gains for learners in an AI-supported condition than for a traditional comparison group. Thongsan and Anderson (2025) similarly found that university EFL learners moved from surface clarification toward more evaluative questioning when ChatGPT was integrated into explicit critical-reading instruction. These studies suggest that AI can support critical engagement when it is embedded in tasks that require learners to interrogate, compare and justify rather than merely retrieve answers.

The same literature also cautions against cognitive offloading. Qian's (2025) systematic review highlights the tension between efficiency and the outsourcing of critical and metacognitive work, while Xu et al. (2025) show that generative-AI environments require explicit metacognitive support if learners are to sustain self-regulation. The present study extends this debate by shifting the unit of analysis from students' direct interaction with an AI system to the instructional text itself. Its central contribution is therefore to investigate whether AI-generated reading passages can serve as transitional scaffolds within a teacher-mediated EFL reading sequence while authentic materials continue to provide complexity and transfer.

Research gap and questions

The emerging literature remains dominated by studies of writing, problem solving, attitudes toward AI, or direct student use of chatbots. Fewer studies isolate AI-generated reading texts as pedagogical materials, fewer still adopt longitudinal comparative designs, and evidence from Moroccan EFL contexts remains scarce. The

current study addresses this gap by comparing critical-thinking development under AI-generated and traditional authentic texts across two educational settings in Tangier.

RQ1. What is the impact of AI-generated reading texts on students' critical-thinking development?

RQ2. How does performance under AI-generated texts compare with performance under traditional authentic texts?

RQ3. Are similar differences observable across CPGE and EMSI contexts?

METHODOLOGY

Research design and setting

The study adopted a comparative longitudinal design across an eight-week instructional period. One condition used traditional authentic argumentative texts from sources such as The Economist and The New York Times, while the other used AI-generated texts designed to match the themes and approximate difficulty level while offering clearer organisation and simplified structure. Performance was documented at Week 1, Week 4 and Week 8. The study was situated in CPGE and EMSI classes in Tangier, both taught by the first author, which allowed the research problem and the intervention to emerge from regular classroom practice.

Participants and instructional context

The study involved students enrolled in the first author's regular English-teaching contexts at CPGE and EMSI in Tangier. The comparative design was organised around two instructional conditions, traditional authentic texts and AI-generated reading texts, and around the two institutional settings. The analysis therefore examined development by text condition and by educational context across the eight-week intervention.

Study characteristic	Documented information
Study sites	CPGE and EMSI, Tangier, Morocco
Instructional context	Regular English classes taught by the first author
Design	Classroom-based comparative longitudinal design
Duration	8 weeks; measurements reported at Weeks 1, 4, and 8
Conditions	Traditional authentic texts vs. AI-generated reading texts
Baseline aggregate scores	Traditional = 55; AI-generated = 54

At baseline, the two instructional conditions showed closely comparable aggregate scores: 55 for the traditional-text condition and 54 for the AI-generated-text condition. This near-equivalence provided a useful starting point for examining how the two groups developed across Weeks 4 and 8 within the normal classroom setting.

Reading materials and intervention

Traditional texts were authentic, complex and argumentative and were drawn from sources such as The Economist and The New York Times. AI-generated materials were designed around comparable themes and an intended level of difficulty, with clearer organisation and a simplified structure. The pedagogical contrast therefore concerned textual accessibility rather than topic alone. In both conditions, reading work focused on the same higher-order objectives of analysis, inference and evaluation.

Instruments and operationalization of critical thinking

The study used pre-test/post-test performance, reading tasks centered on analysis, inference and evaluation, a student questionnaire on perceptions and engagement, and classroom observation of participation and interaction. The performance measure was a study-specific classroom assessment aligned with the three focal critical-thinking dimensions. These dimensions correspond closely to the Delphi framework articulated by Facione (1990), providing a clear conceptual link between the instructional objectives and the assessment focus.

Measurement validity and interpretive consistency

Measurement credibility was supported through close alignment between the construct under investigation, the instructional tasks and the assessment dimensions. Analysis, inference and evaluation were targeted consistently throughout the intervention and were also used to organise the reading tasks and performance comparisons. This alignment strengthens the content validity of the classroom measure and supports its use for examining longitudinal change across the two text conditions. The scores are therefore interpreted as study-specific indicators of critical-reading performance rather than as a broad general measure of all dimensions of critical thinking.

Measurement property	Documented/interpretive status
Construct focus	Analysis, inference, and evaluation aligned with established critical-thinking frameworks
Instrument type	Study-specific classroom performance measure
Content validity	Assessment dimensions aligned with the intervention objectives and reading tasks
Interpretive scope	Scores used to compare critical-reading performance across time and text conditions

Questionnaire and classroom observation

The questionnaire and observation components served as supplementary process measures. The questionnaire focused on student perceptions and engagement, while classroom observation focused on participation and interaction patterns. Together, these sources provided contextual evidence for interpreting the performance results and for examining how students responded to the two types of reading material in regular classroom practice.

Statistical analysis

The statistical analysis focused on longitudinal descriptive comparison across the two text conditions and the two institutional contexts. The analysis reports Week 1, Week 4 and Week 8 scores, absolute gains, relative gains, the changing between-condition gap, a descriptive difference-in-differences, and a gain ratio. Relative gain was calculated as $(\text{post-test} - \text{pre-test}) / \text{pre-test} \times 100$. The descriptive difference-in-differences was calculated as $(82 - 54) - (69 - 55) = 14$ points. These indices provide a more precise picture of the magnitude and direction of change across the eight-week period.

Threats to internal validity and researcher role

Several validity considerations remain relevant in a classroom-based longitudinal study. Repeated testing across eight weeks may contribute to practice effects, while normal academic development may contribute to maturation. In addition, the first author taught in both institutional settings. This sustained classroom access strengthened ecological validity and contextual understanding, while also making teacher mediation an integral part of the intervention itself. The findings are therefore interpreted within the naturally occurring instructional contexts in which the study was conducted.

RESULTS

Descriptive longitudinal development

At Week 1, the two conditions were nearly equivalent at the aggregate level: 55 for the traditional-text condition and 54 for the AI-generated-text condition. By Week 4, scores had risen to 61 and 66 respectively. At Week 8, the traditional condition reached 69, whereas the AI condition reached 82. Thus, the traditional condition gained 14 points and the AI condition gained 28 points across the eight-week period.

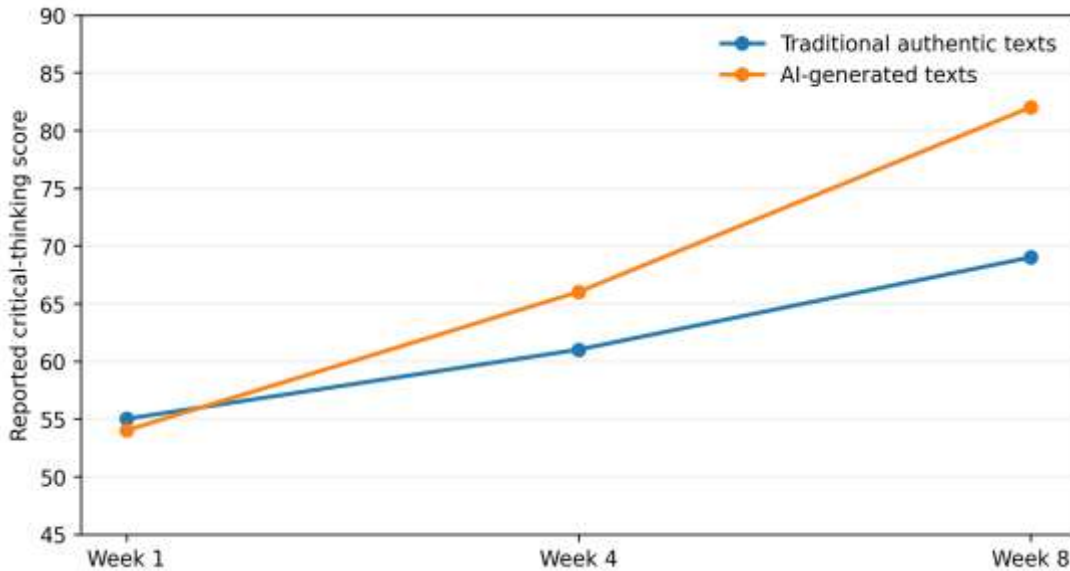


Figure 1. Critical-thinking performance across the eight-week study

Note. Values are the study's reported aggregate scores.

The between-condition gap changed from -1 point for the AI condition at baseline to $+5$ points at Week 4 and $+13$ points at Week 8. The descriptive difference-in-differences is therefore 14 points: the AI-text condition improved by 14 points more than the traditional-text condition over the same period. This widening gap is consistent with a progressively stronger performance trajectory under the AI-generated-text condition.

Table 1. Descriptive longitudinal indices by text condition

Index	Traditional	AI-generated	AI – Traditional
Week 1 score	55	54	-1
Week 4 score	61	66	$+5$
Week 8 score	69	82	$+13$
Absolute gain	$+14$	$+28$	$+14$
Relative gain	25.5%	51.9%	$+26.4$ percentage points

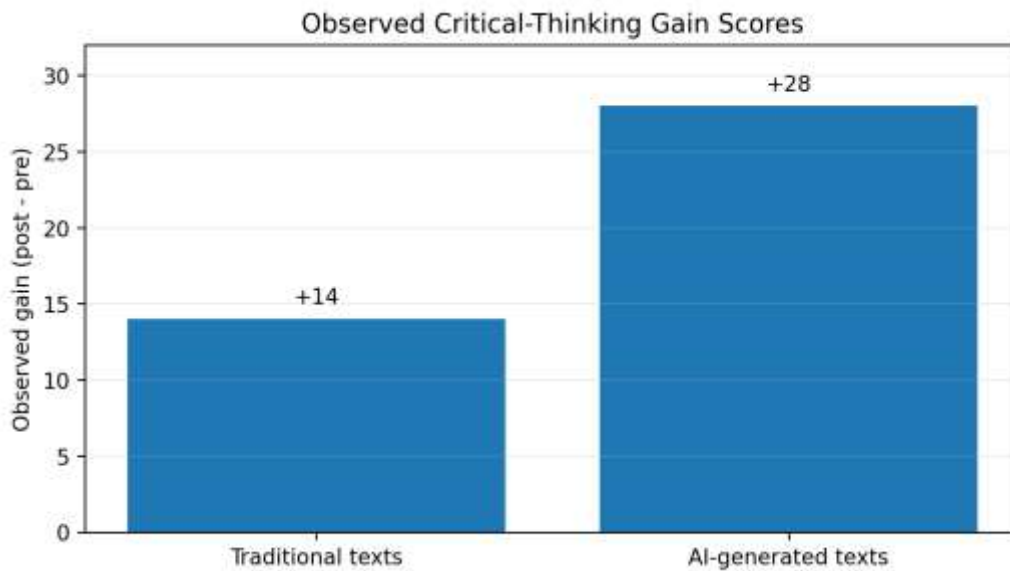


Figure 4. Observed gain scores by text condition

Note. The AI-generated-text condition improved by 28 points, exactly twice the 14-point observed gain in the traditional-text condition. These are descriptive gains, not standardised effect sizes.

Expressed relative to baseline, the traditional condition improved by 25.5%, whereas the AI-generated condition improved by 51.9%. The observed gain ratio is therefore 2.00, meaning that the AI-text condition recorded twice the absolute improvement observed under traditional texts. Taken together with the longitudinal pattern, these indices provide a clearer picture of the magnitude of the recorded difference.

Pre-test/post-test comparison

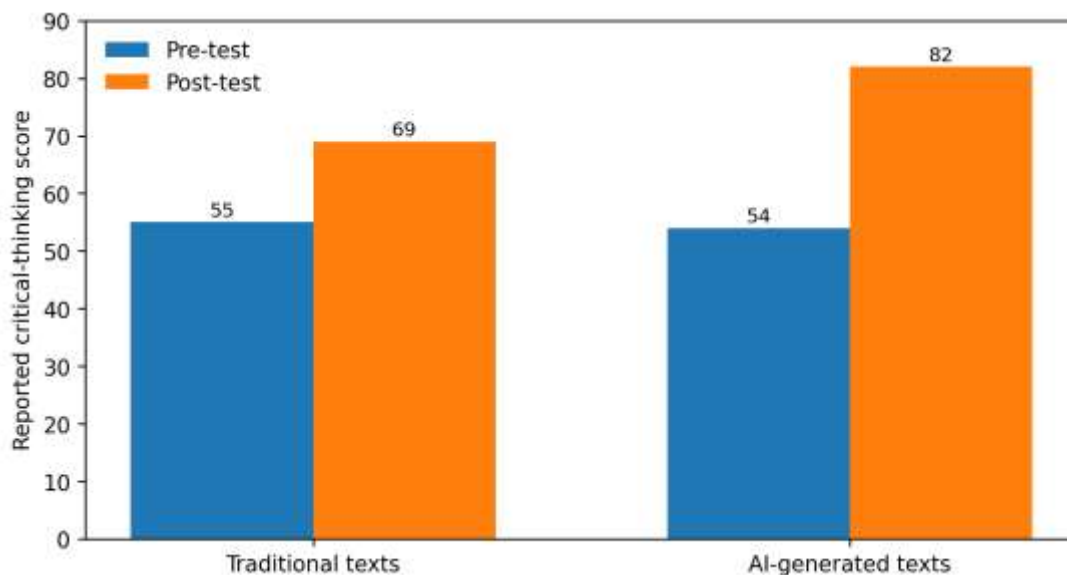


Figure 2. Pre-test and post-test performance by text condition

Note. Traditional: 55 to 69 (+14). AI-generated: 54 to 82 (+28).

The pre-test/post-test pattern is notable because the AI condition began one point below the traditional condition but finished 13 points above it. This reversal is consistent with accelerated improvement under the AI-generated materials and reinforces the longitudinal pattern already visible at the midpoint of the intervention.

CPGE and EMSI comparison

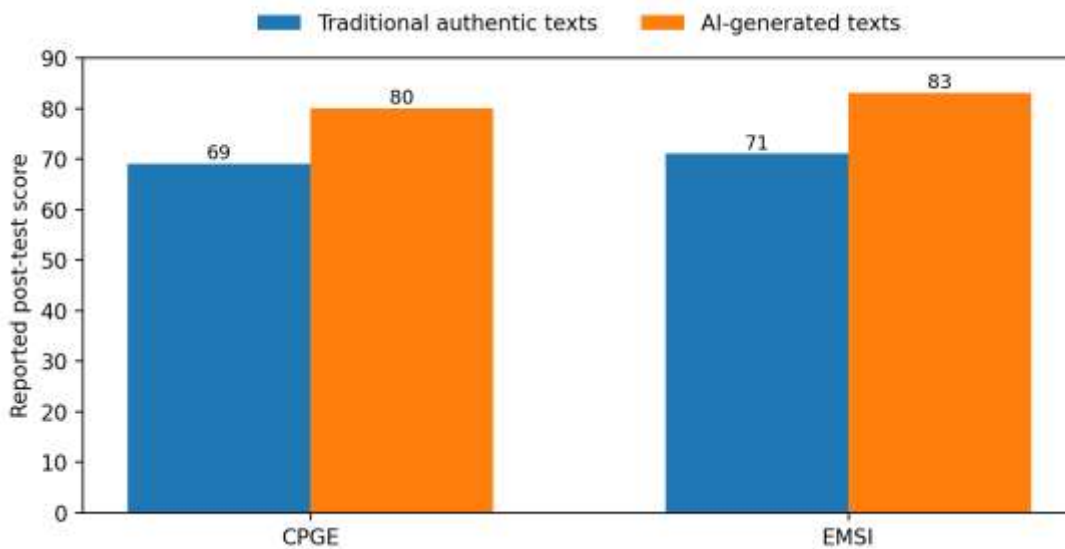


Figure 3. Reported post-test performance across CPGE and EMSI

Note. CPGE: 69 traditional, 80 AI-generated. EMSI: 71 traditional, 83 AI-generated.

In CPGE, the AI-text condition was 11 points higher than the traditional-text condition at the reported final comparison; in EMSI, the difference was 12 points. This cross-context consistency is one of the study's principal strengths because the same directional pattern appeared in two distinct educational environments. The near-parallel differences also suggest that the pedagogical value of the AI-generated-text condition was not confined to a single institutional setting.

Questionnaire and classroom-observation results

The questionnaire and classroom-observation components complemented the performance data by focusing on student perceptions, engagement, participation and interaction. The classroom-level pattern indicated that AI-generated texts improved accessibility, reduced processing difficulty and supported engagement, whereas traditional texts preserved greater textual complexity and encouraged deeper evaluative work. These process-oriented observations help contextualise the numerical results and reinforce the study's interpretation of AI-generated texts as scaffolding rather than replacement.

Supplementary component	Documented status/findings
Questionnaire focus	Student perceptions and engagement
Observation focus	Participation and interaction patterns
AI-generated-text pattern	Greater accessibility, lower processing difficulty, and increased engagement
Traditional-text pattern	Greater textual complexity and deeper evaluation demands
Analytic role	Contextual process evidence used alongside the longitudinal performance results

Summary of the statistical pattern

Across the eight-week period, the AI-generated-text condition moved from a one-point baseline disadvantage to a 13-point final advantage. Its absolute gain was 28 points, compared with 14 points in the traditional condition, while its relative gain was 51.9% compared with 25.5%. The resulting gain ratio of 2.00 and the 14-

point descriptive difference-in-differences both point to a substantially steeper trajectory under the AI-generated-text condition. The CPGE and EMSI comparisons reproduce the same direction of difference, adding contextual consistency to the overall pattern.

DISCUSSION

AI-generated texts as scaffolding rather than replacement

The documented pattern supports the study's central pedagogical proposition: AI-generated reading texts may be useful as scaffolding when they reduce unnecessary linguistic and organisational barriers while leaving the higher-order work of analysis, inference and evaluation to the learner. The AI condition's larger observed gain is compatible with cognitive-load theory, according to which reducing extraneous processing can free resources for learning-relevant cognition (Sweller, 1988). This interpretation also converges with Liu and Wang (2024) and Thongsan and Anderson (2025), who report stronger critical engagement when AI-supported activities are embedded in explicit pedagogical design.

At the same time, the findings do not justify replacing authentic texts. Authentic newspaper and magazine texts preserve rhetorical complexity, ambiguity and naturally occurring argumentative structure. These features are precisely what advanced readers eventually need to negotiate. A defensible pedagogical sequence is therefore progressive: AI-generated texts can provide an accessible entry into demanding themes and reasoning tasks, after which authentic materials test transfer and independent critical reading. This preserves the study's balanced perspective on AI as scaffolding rather than replacement.

Dependence, metacognition, and teacher mediation

The study's assistance-dependence framing becomes especially important when the findings are interpreted alongside recent research. Qian (2025) identifies cognitive offloading as a recurring risk in higher education, and Xu et al. (2025) show that learners in generative-AI environments benefit from explicit metacognitive support. The pedagogical issue is therefore not simply whether AI makes reading easier. It is whether the reduction in difficulty is strategically targeted. If AI removes lexical or organisational obstacles while teachers require students to justify interpretations, compare evidence and challenge claims, assistance can support learning. If AI supplies the interpretations themselves, assistance may become dependence.

Contribution to Moroccan EFL research

The study contributes to the literature by examining a form of AI integration that has received less attention: teacher-designed AI-generated reading texts rather than student chatbot use. Its longitudinal comparison, its inclusion of both CPGE and EMSI, and its focus on Moroccan EFL reading distinguish it from studies centred primarily on writing, general AI attitudes or direct prompting. The cross-context design is especially useful because CPGE and engineering students encounter demanding academic reading for different educational purposes, yet the same descriptive pattern appeared in both settings. The contribution is therefore contextual as well as pedagogical: it provides preliminary evidence from a setting that remains underrepresented in the rapidly expanding generative-AI literature.

Cross-context consistency and the limits of generalisation

The 11-point CPGE difference and 12-point EMSI difference are descriptively similar, giving the study a degree of contextual robustness. The two institutions nevertheless differ in programme structure, assessment culture and students' academic trajectories. The importance of the cross-context comparison therefore lies in the recurrence of the same pedagogical pattern across two demanding Moroccan post-secondary environments rather than in treating those environments as interchangeable.

Pedagogical Implications

The findings favour a staged dual-text model. Teachers may begin with AI-generated texts calibrated for clarity and level, use questions that explicitly target analysis, inference and evaluation, and then move students toward authentic materials on comparable themes. A second implication concerns task design: the cognitive demand should reside in what students do with the text, not merely in how difficult the sentences are to decode. Comparison tasks, counterargument generation, evidence ranking and bias detection can preserve intellectual challenge even when the language has been scaffolded.

AI literacy should also be integrated into critical reading. Learners can compare an AI-generated passage with an authentic source, identify omitted qualifications or shifts in emphasis, verify claims and discuss how simplification changes meaning. Such practices maintain human judgment at the centre of the activity and directly address the risk of dependence identified in recent research.

Limitations and Future Research

The study should be interpreted within the scope of a classroom-based longitudinal comparison conducted in two specific Moroccan post-secondary settings. Its use of aggregate group-level outcomes provides a clear picture of directional change but places greater emphasis on group trajectories than on individual variability. A second consideration concerns measurement: the critical-thinking assessment was designed around the study's focal dimensions of analysis, inference and evaluation. Future studies could strengthen this line of inquiry further by combining such classroom performance measures with standardised instruments and formal psychometric evaluation.

The questionnaire and classroom-observation components enriched the contextual interpretation of the intervention by focusing on perceptions, engagement, participation and interaction. Future research could extend this component through more systematic longitudinal coding and closer integration between process measures and performance outcomes. Most importantly, subsequent studies should incorporate a delayed post-test using fully authentic materials after the AI scaffold has been removed. Such a design would test whether gains developed under AI-generated scaffolding transfer to later independent reading and would address the central risk of dependence directly.

CONCLUSION

The eight-week descriptive record shows improvement under both text conditions, with a steeper trajectory under AI-generated reading texts. Scores in the AI condition rose from 54 to 82, compared with 55 to 69 in the traditional condition, and the same direction of difference appeared in CPGE and EMSI. These results are consistent with the view that carefully designed AI-generated texts can function as cognitive scaffolds when they are embedded in teacher-mediated tasks that preserve analysis, inference and evaluation.

As a classroom-based longitudinal comparison, the study provides a coherent pattern of development across two text conditions and two demanding Moroccan educational contexts. Its value lies in showing that AI-generated texts can support stronger observed growth when they are used as transitional scaffolds within teacher-mediated critical-reading instruction. The central pedagogical message remains clear: AI is most useful when it supports learners on the way to independent engagement with authentic complexity, not when it replaces that engagement.

REFERENCES

1. Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership*, 43(2), 44–48.
2. Essien, A., Bukoye, O. T., O'Dea, X., & Kremantzis, M. (2024). The influence of AI text generators on critical thinking skills in UK business schools. *Studies in Higher Education*, 49(5), 865–882. <https://doi.org/10.1080/03075079.2024.2316881>

3. Facione, P. A. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. American Philosophical Association.
4. Grabe, W. (2009). Reading in a second language: Moving from theory to practice. Cambridge University Press.
5. Kintsch, W. (1998). Comprehension: A paradigm for cognition. Cambridge University Press.
6. Kasneci, E., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
7. Liu, W., & Wang, Y. (2024). The effects of using AI tools on critical thinking in English literature classes among EFL learners: An intervention study. *European Journal of Education*, 59(4), e12804. <https://doi.org/10.1111/ejed.12804>
8. Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. *TechTrends*, 69, 1105–1120. <https://doi.org/10.1007/s11528-025-01100-1>
9. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
10. Thongsan, N. C., & Anderson, N. J. (2025). From passive answers to active inquiry: How AI supports critical reading in EFL classrooms. *LEARN Journal: Language Education and Acquisition Research Network*, 18(2), 795–820. <https://doi.org/10.70730/KMKL8505>
11. Xu, X., Qiao, L., Cheng, N., Liu, H., & Zhao, W. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *British Journal of Educational Technology*, 56, 1842–1863. <https://doi.org/10.1111/bjet.13599>