Article

Journal of Interdisciplinary Studies in Education

Volume 15, Issue 8 (2026)
Journal of Interdisciplinary Studies in Education
ISSN: 2166-2681Print 2690-0408 Online | https://ojed.org/jise
https://doi.org/10.32674/enffa425

Beyond Comprehension: Assessing Moroccan CPGE First-Year Students' Readiness for AI-Era Critical Literacy Demands

Zakaria El Hajji
Chouaib Doukkali University, El Jadida, Morocco
ORCID: 0009-0002-5012-7563

Nadia Hellalet
Chouaib Doukkali University, El Jadida, Morocco
ORCID: 0000-0003-1879-8164

ABSTRACT

The rapid proliferation of generative artificial intelligence in educational contexts has transformed how students encounter academic content, yet a fundamental question remains underexamined: Do English as a Foreign Language (EFL) learners possess the critical literacy competencies needed to evaluate AI-generated texts? This study investigated whether students in Morocco's Classes Préparatoires aux Grandes Écoles (CPGE)—one of the nation's most academically selective post-secondary programs—demonstrate readiness for AI-era literacy demands. We developed a diagnostic assessment instrument, the Critical Reading Diagnostic Assessment (CRDA), designed to mirror common characteristics of AI-generated academic content: a confident, authoritative tone, a logical structure with subtle gaps, claims requiring verification, and culturally embedded assumptions. Thirty first-year CPGE students (M age = 18.6 years) evaluated argumentative passages across five critical dimensions: argument identification, evidence evaluation, assumption recognition, multi-source synthesis, and source criticism. Metacognitive self-assessment data were also collected. Results revealed systematic evaluation failures across all dimensions, with mean accuracy rates of only 18–39% and large effect sizes confirming substantial deficits (d = 1.11–2.25). Metacognitive analysis showed that students overestimated their abilities by 11–38 percentage points across all dimensions, indicating profound self-assessment inaccuracies. These findings indicate that Moroccan EFL students are fundamentally unprepared for AI literacy demands and argue for a diagnostic-first approach to AI integration: assess critical evaluation competencies, provide targeted instruction, then introduce AI tools with explicit evaluation frameworks.

Keywords: AI literacy, ChatGPT, critical evaluation, diagnostic assessment, EFL learners, generative AI, Morocco, metacognition

© Author(s), 2026. Published by Star Scholars Press.
This article is distributed under the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. https://creativecommons.org/licenses/by/4.0/

INTRODUCTION

Picture a first-year CPGE classroom in Morocco where students encounter an argument about renewable energy policy from a well-established news outlet. The passage appears authoritative—confidently articulated claims, specific statistics, and citations to peer-reviewed research. Students engage carefully: they identify unfamiliar vocabulary, parse complex syntactic structures, and extract main points. When asked to summarize, they succeed. But there is something these students do not know: the passage was deliberately designed to mirror characteristics common in AI-generated academic content. Embedded within its fluent prose are a fabricated study citation, a logical gap masked by an authoritative tone, and an unstated assumption reflecting culturally situated perspectives. When asked to evaluate the quality of the evidence, identify underlying assumptions, and assess source credibility, most students fail. They accept the problematic content uncritically, unable to distinguish confident presentation from credible argumentation.

This scenario is not hypothetical. It describes the present study and highlights a crisis emerging at the intersection of artificial intelligence proliferation and educational literacy. As generative AI tools like ChatGPT become ubiquitous in teaching and learning—with teachers using AI to generate instructional materials, students using AI for research and writing assistance, and institutions investing in AI-powered educational platforms—a fundamental question remains underexamined: Do students possess the critical literacy competencies needed to engage safely and effectively with AI-generated content?

This question is urgent for Morocco's Classes Préparatoires aux Grandes Écoles (CPGE) first-year students. These programs are the nation's most academically selective post-secondary pathway, admitting approximately the top 10% of national baccalaureate graduates. CPGE students demonstrate exceptional academic performance and face rigorous curricula designed to prepare them for the French grandes écoles system. If these students lack the critical literacy skills necessary for AI-era academic engagement, the implications for broader student populations are profound.

Morocco has invested substantially in educational transformation. The 2015 Strategic Vision for Reform explicitly prioritizes "critical thinking, creativity, and problem-solving" as foundational competencies essential for 21st-century success (Conseil Supérieur de l'Éducation, 2015, p. 12). Yet traditional instructional approaches, marked by teacher-centered delivery and assessments focused on factual recall, continue to dominate many Moroccan classrooms (Amrous & Nejmaoui, 2016). These pedagogies cultivate important foundations—disciplined study habits, content knowledge, linguistic competence—but may offer limited opportunities to develop the evaluative skepticism and analytical independence that critical literacy requires. Now, as generative AI introduces new complications, these pedagogical limitations take on added significance.

Existing research on critical literacy in EFL contexts documents serious deficiencies but leaves crucial gaps unaddressed. Studies have examined what students write rather than how they evaluate the arguments they encounter—a distinction that is especially critical in the AI era, when students increasingly consume AI-generated content that requires evaluation. Moreover, AI-specific evaluation challenges—hallucinations, training biases, and surface coherence masking logical gaps—compound existing literacy demands in ways that have not yet been examined empirically, particularly in non-Western, multilingual educational contexts.

We address three research questions. First, what are Moroccan CPGE first-year students' actual competencies across five critical evaluation dimensions essential for assessing AI-generated content: argument identification, evidence evaluation, assumption recognition, multi-source synthesis, and source criticism? Second, which sub-skills show the most severe deficits and require prioritized instructional attention? Third, do students accurately recognize their own evaluation limitations, or do metacognitive gaps prevent them from judging when AI-generated content has misled them?

LITERATURE REVIEW

Theoretical Frameworks for Critical Reading and Evaluation

Scholars conceptualize critical reading as active, evaluative engagement with texts that extends beyond literal comprehension to assessment of argumentation quality, evidence sufficiency, and rhetorical purpose. Kurland (2000) characterizes it as "an active process of discovery" wherein readers interrogate rather than passively absorb content, evaluating reasoning, assessing evidence, identifying bias, recognizing unstated assumptions, and considering alternative interpretations. This emphasis on active interrogation distinguishes critical reading from comprehension: students may successfully extract a text's explicit meaning while failing to evaluate whether that meaning rests on sound reasoning or credible evidence.

Paul and Elder (2006) embed critical reading within their broader critical thinking framework, arguing that proficient readers apply consistent intellectual standards—clarity, accuracy, precision, relevance, depth, breadth, logic, fairness—to textual claims. Crucially, they emphasize metacognitive awareness: expert readers monitor their own thinking processes, recognize when biases influence interpretation, and consciously work to overcome cognitive limitations. This metacognitive dimension is important for AI literacy; students who cannot accurately assess their own evaluation competencies may overestimate their ability to judge AI-generated content quality.

Grabe and Stoller (2020) decompose critical reading into specific component skills: recognizing argument structures and distinguishing main claims from supporting premises, evaluating whether evidence supports assertions made, detecting bias, identifying purposes beyond surface claims, recognizing unstated assumptions, assessing source credibility, and synthesizing information across multiple texts. This componential approach enables diagnostic assessment—identifying which specific skills students lack rather than treating critical reading as an undifferentiated competency. Pressley and Afflerbach (1995) add that skilled readers engage in continuous self-monitoring and strategic deployment of repair mechanisms, and that these metacognitive processes distinguish experts from novices more than domain knowledge or reading speed.

Critical Literacy in EFL Contexts: Evidence and Challenges

Research on critical thinking and literacy in EFL contexts highlights challenges that are particularly evident in settings where English is students' third or fourth language and where educational traditions emphasize knowledge transmission over critical questioning. Hellalet (2021) found that Moroccan university students "exhibited a low level of critical thinking skills," struggling to "develop a stand and support it with arguments and counter-arguments" (p. 179). Hellalet attributed these difficulties, in part, to secondary education that rewards memorization rather than questioning. Amrous and Nejmaoui (2016) identified systemic pedagogical factors, noting that instruction typically prioritizes linguistic form over analytical substance.

Extending beyond Morocco, research across MENA EFL contexts confirms similar patterns. Alwadai (2014) found that Saudi teachers lacked the knowledge and capacity to teach critical thinking effectively, documenting systemic obstacles such as weak pre-service preparation and classroom structures that inhibit evaluative engagement. Atkinson (1997) questioned whether critical thinking, as conceptualized in Western educational traditions, transfers seamlessly to non-Western contexts, suggesting that cultural and educational traditions may value different epistemic approaches. However, subsequent research has challenged this cultural essentialism, demonstrating that, with appropriate instruction, students from diverse cultural backgrounds develop critical literacy effectively (Manalo & Sheppard, 2016).

Research in comparable EFL contexts supports this pedagogical explanation. Thongsan and Anderson (2025) examined the effects of AI-supported reading instruction on critical reading skills among Thai university EFL students and found that, even with AI scaffolding, learners continued to struggle to evaluate arguments and recognize bias in texts. The authors concluded that AI tools, when used without explicit instruction in critical literacy, function as comprehension aids rather than as evaluation trainers.

Three patterns from this literature are relevant here. First, critical literacy challenges in EFL contexts appear attributable more to pedagogical approaches than to inherent cultural incompatibilities. Second, the linguistic demands of reading in a non-native language compound the cognitive load required for critical evaluation. Third, existing research focuses predominantly on production rather than evaluation, creating a gap that is especially salient in AI literacy contexts, where students primarily consume AI-generated content.

Generative AI and New Literacy Demands

The emergence of large language models like ChatGPT has created literacy demands that build on but extend beyond traditional critical reading competencies. First, AI systems produce "hallucinations"—confident assertions of false information presented without uncertainty markers (Ji et al., 2023). Unlike human authors who typically hedge uncertain claims, language models generate all content with the same level of confidence, regardless of veracity, because they rely on statistical pattern matching that cannot distinguish verified facts from plausible-sounding fabrications (Bender et al., 2021). Research documents numerous examples: ChatGPT has been observed citing nonexistent academic studies, fabricating statistical data, and asserting unsupported causal relationships while maintaining an authoritative tone throughout. Darvin (2025) extends this concern by arguing that GenAI platforms trained on large datasets privilege certain ways of thinking and writing. This foregrounds critical digital literacy as an essential competency for L2 learners navigating AI-mediated contexts.

Second, AI inherits biases embedded in training data that predominantly reflect Western, Anglophone, economically developed societies (Bender et al., 2021). These systems make unstated assumptions about "normal" educational systems, "typical" family structures, and "standard" economic development pathways—assumptions that may contradict realities in multilingual, postcolonial, and economically diverse contexts. For Moroccan students navigating among Tamazight, Darija, Modern Standard Arabic, French, and English, AI's predominantly English-language training makes it harder to recognize when outputs embed culturally specific assumptions rather than universal truths.

Third, AI-generated texts exhibit surface-level coherence that can mask logical gaps or reasoning flaws (Wang et al., 2023). Appropriate transitions, consistent structure, grammatical accuracy, and domain-appropriate vocabulary signal credibility and may suppress critical evaluation. Research in cognitive psychology shows that processing fluency shapes credibility judgments: people are more likely to believe claims presented in clear, fluent language than identical claims presented less fluently (Alter & Oppenheimer, 2009), a pattern especially salient for EFL learners who may equate linguistic sophistication with intellectual authority.

Despite AI's rapid integration into educational contexts, research on students' AI literacy competencies remains limited. Recent studies have examined AI's potential to support language learning (Xiao et al., 2025), generate writing feedback (Ray, 2023), and develop information literacy skills in AI contexts (Buselic, 2023), yet investigations into whether students can critically evaluate AI-generated content are scarce. Existing studies focus primarily on university students in Western contexts, leaving questions about AI literacy in multilingual, non-Western settings largely unexplored. Even when aware of AI's origin, students often fail to verify AI claims or question AI assertions (Xiao et al., 2025). Wang and Wang (2025) propose an APSE (Awareness, Practice, Strategy, Evaluation) model for critical AI literacy, identifying evaluation as the most underdeveloped competency among L2 learners—a finding that directly anticipates the deficits documented in the present study. If students lack critical evaluation competencies for human-authored texts, AI's authoritative tone and surface coherence may make evaluation more difficult rather than less. Sardi et al. (2025) find in a systematic review that students who rely on AI without evaluative frameworks risk significant erosion of independent analytical capacity.

Summary and Research Gap

The literature establishes that critical reading comprises specific, assessable competencies, including argument identification, evidence evaluation, assumption recognition, source criticism, and multi-source synthesis. EFL learners, particularly in MENA contexts, often struggle with these competencies, though evidence suggests that pedagogical approaches, rather than cultural incompatibilities, primarily drive these difficulties. Emerging research on generative AI shows that large language models produce content with characteristics—confident hallucinations, embedded training biases, and surface coherence that mask logical gaps—that compound existing evaluation challenges. Melisa et al. (2025) confirm that AI proliferation in higher education has intensified rather than resolved the need for critical thinking instruction, lending urgency to the three remaining gaps: we lack diagnostic data on whether EFL students possess the specific critical evaluation competencies necessary to engage with AI-generated content; existing research focuses on Western contexts; and studies have not investigated metacognitive dimensions—whether students recognize their own evaluation limitations or overestimate their capacity to judge AI content quality. The present study addresses these gaps.

METHOD

Participants

Participants comprised 30 first-year students (18 female, 12 male; *M* age = 18.6 years, *SD* = 0.8, range = 18–20 years) enrolled in the scientific track of a CPGE program in Mohammedia, Morocco. CPGE programs are Morocco's most academically selective post-secondary pathway, admitting approximately the top 10% of national baccalaureate graduates based on examination grades. All participants had completed Moroccan secondary education, where instruction occurs primarily in Arabic and French, with English taught as a foreign language beginning in middle school. At the time of data collection (September 2025), participants were in the first week of their CPGE program. English proficiency levels, assessed through program entrance examinations, ranged from A2 to C1 on the CEFR, with most students (n = 23, 77%) scoring at the A2 level.

Sampling used purposive selection to focus on first-year students during their initial CPGE trimester, before extensive exposure to the program's academic reading demands could influence baseline critical evaluation competencies. The sample size (N = 30) was determined by previous validation studies of critical reading instruments in EFL contexts (Liu & Stapleton, 2014) and by practical constraints in accessing this highly selective student population. The study received institutional ethics approval, and all participants provided informed consent before data collection.

Instruments

Critical Reading Diagnostic Assessment (CRDA)

To assess students' readiness for AI-era literacy demands, we designed the CRDA to mirror characteristics commonly observed in AI-generated academic content, serving dual purposes: measuring critical evaluation competencies relevant to any complex argumentative text and providing ecological validity for the AI-saturated contexts students increasingly navigate. The CRDA incorporated five distinctive features identified in research on large language model outputs (Bender et al., 2021; Ji et al., 2023; Wang et al., 2023): (1) a confident, authoritative tone without hedging or uncertainty markers; (2) a logical structure with subtle gaps that mask logical leaps; (3) claims requiring verification, including specific statistics and study citations; (4) unstated assumptions reflecting particular worldviews; and (5) evidence requiring evaluation for relevance and sufficiency.

We designed assessment passages rather than using raw AI outputs to ensure precise control over difficulty levels, linguistic complexity appropriate for A2–C1 proficiency students, and systematic

alignment with our five assessment dimensions. The final CRDA comprised four sections targeting authentic text types students increasingly encounter in AI-mediated academic contexts. Section 1 collected metacognitive self-assessments across the five evaluation dimensions using 5-point Likert scales. Section 2 (Text A) presented an approximately 280-word argumentative opinion piece on digital classroom technology, accompanied by seven items assessing argument identification, evidence evaluation, assumption recognition, and critical stance. Section 3 (Text B) presented a bar chart infographic on young adults' news consumption habits, accompanied by four items targeting data interpretation, headline evaluation, and source criticism. Section 4 (Text C) presented a social media promotional post, accompanied by three items assessing source credibility, identification of missing evidence, and recognition of persuasive intent. Section 5 included one general critical literacy item assessing students' self-reported epistemic response to contradictory information. This multi-text design reflects the diversity of text types students encounter when navigating AI-generated and digitally mediated content, providing ecological validity beyond single-genre assessment.

Instrument validation

Three faculty members with expertise in critical thinking pedagogy and EFL assessment reviewed the CRDA for content validity, providing feedback on item clarity, construct alignment, and cultural appropriateness. Based on their recommendations, we revised four items for greater precision and added examples to two complex items. To establish inter-rater reliability, two trained raters independently scored a randomly selected subset (n = 10, 33%) of completed assessments. Cohen's kappa indicated substantial agreement (κ = .87, p < .001). Internal consistency for the full instrument was acceptable (Cronbach's α = .78).

Metacognitive self-assessment

Before completing the CRDA, students rated their competencies across the five evaluation dimensions on 5-point Likert scales (1 = very poor, 5 = excellent), with brief descriptions of each dimension adapted for B2-level English proficiency. This allowed us to calculate metacognitive accuracy by comparing self-rated competency with actual performance.

Procedure

Data collection occurred during a single 90-minute session in September 2025 in participants' regular classroom setting. The research team explained the study purpose, obtained informed consent, and assured participants that performance would not affect their grades. Students completed the metacognitive self-assessment (approximately 10 minutes), followed by the CRDA (60–70 minutes). Participants could use dictionaries but not collaborate or access internet resources. After completing the CRDA, we conducted brief semi-structured interviews with a purposive subsample (n = 4) selected to represent performance variation. These qualitative data provided interpretive context but are not reported in detail due to space constraints. Interview data are reserved for a forthcoming qualitative analysis.

Data Analysis

We analyzed data in SPSS Version 28.0. Descriptive statistics (means, standard deviations, frequency distributions) characterized performance across the five evaluation dimensions. To assess whether performance differed significantly from chance (25%), we conducted one-sample t-tests. Effect sizes (Cohen's d) quantified the magnitude of differences. For metacognitive accuracy, we converted self-ratings to percentage scales (1 = 0%, 5 = 100%) to enable direct comparison with performance percentages. Metacognitive gap scores were calculated by subtracting actual performance from self-rated ability; positive gaps indicate overestimation. Paired-samples t-tests assessed whether gaps differed significantly

from zero. We report exact p-values and effect sizes to enable assessment of both statistical and practical significance. All tests used two-tailed criteria with $\alpha = .05$.

RESULTS

This section presents findings addressing three research questions: (1) What are students' actual competencies across five critical evaluation dimensions? (2) Where do students show the most severe deficits? (3) Do students accurately recognize their evaluation limitations?

Overall Performance Across Dimensions

Table 1 presents descriptive statistics and inferential test results for performance across five critical evaluation dimensions. Students consistently performed at low levels, with mean accuracy ranging from 18.3% (assumption recognition) to 38.5% (multi-source synthesis). One-sample t-tests compared performance against a 25% chance benchmark; none of the five dimensions reached statistical significance (all $p > .05$), indicating that student performance was statistically indistinguishable from random responding across all dimensions. Effect sizes were negligible to small ($d = -0.17$ to 0.28), confirming the absence of meaningful deviation from chance in either direction. The proportion of students failing to reach minimum proficiency (≥50% accuracy) ranged from 73.3% (argument identification) to 96.7% (assumption recognition), providing the most practically meaningful indicator of the depth of the deficit.

Table 1

Performance Across Five Critical Evaluation Dimensions (N = 30)

Dimension	M (%)	SD	n (%) ≥ 50%	t(29)	p	Cohen's d
Argument Identification	26.7	44.6	8 (26.7%)	0.21	.836	0.04
Evidence Evaluation	34.3	47.1	4 (13.3%)	1.08	.289	0.20
Assumption Recognition	18.3	39.0	1 (3.3%)	−0.94	.355	−0.17
Multi-Source Synthesis	38.5	48.9	4 (13.3%)	1.51	.142	0.28
Source Criticism	28.9	45.8	3 (10.0%)	0.43	.671	0.08

Note. One-sample t-tests compared mean accuracy on each dimension against a chance benchmark of 25% (four-option items). Cohen's d effect sizes reflect the standardized difference between observed performance and chance: small = 0.20, medium = 0.50, large = 0.80 (Cohen, 1988). All obtained effect sizes fell below the small threshold ($|d| \leq 0.28$), consistent with performance statistically indistinguishable from chance. n (%) ≥ 50% = number and proportion of students meeting the minimum proficiency criterion. Standard deviations reflect high within-group variability on binary and partial-credit items, where individual scores cluster near floor (0%) or ceiling (100%); values exceeding the mean are expected given this bimodal distributional pattern and do not indicate negative scores.

Dimension-Level Findings

Argument identification

Students had difficulty distinguishing main claims from supporting premises. Mean accuracy was 26.7%, marginally above the 25% chance benchmark but not significantly so, $t(29) = 0.21$, $p = .836$, $d = 0.04$. Most students ($n = 22$, 73.3%) scored below 50% proficiency. Common errors included mistaking examples for main claims, confusing premises with conclusions, and identifying topics rather than arguments.

Evidence evaluation

Evidence evaluation proved equally challenging (M = 34.3%), with only 4 students (13.3%) reaching minimum proficiency. Performance did not differ significantly from chance, $t(29) = 1.08$, $p = .289$, $d = 0.20$, indicating that students evaluated evidence quality essentially at random. Common errors included accepting irrelevant evidence as sufficient support and treating anecdotal examples as equivalent to systematic evidence.

Assumption recognition

Assumption recognition yielded the lowest mean accuracy (M = 18.3%), marginally below the 25% chance benchmark. However, the difference was not statistically significant, $t(29) = -0.94$, $p = .355$, $d = -0.17$. Only 1 student (3.3%) achieved minimum proficiency; the remaining 96.7% (n = 29) failed to identify unstated assumptions, instead treating explicitly stated claims as the entirety of the argumentative content. The near-zero success rate on this dimension is the most striking descriptive finding in the dataset. *Multi-source synthesis*. Multi-source synthesis showed the highest mean accuracy (M = 38.5%), though performance remained below minimum proficiency standards and did not differ significantly from chance, $t(29) = 1.51$, $p = .142$, $d = 0.28$. Four students (13.3%) achieved minimum proficiency. Students demonstrated limited capacity to identify contradictions between sources and were largely unable to resolve those contradictions or synthesize coherent understanding across texts.

Source criticism

Source criticism was similarly challenging (M = 28.9%), with only 3 students (10.0%) reaching minimum proficiency. Performance did not differ significantly from chance, $t(29) = 0.43$, $p = .671$, $d = 0.08$. Students rarely verified factual claims, seldom assessed whether cited sources existed, and demonstrated limited capacity to evaluate source credibility or identify conflicts of interest.

The absence of significant deviation from chance across all dimensions is itself a substantive finding: it indicates that students were unable to apply systematic evaluation strategies on any dimension, with responses approximating random selection. The descriptive and proportional data therefore carry greater interpretive weight than the inferential statistics in characterising the nature and depth of these deficits.

Metacognitive Accuracy

Table 2 compares metacognitive self-assessments with actual performance, revealing systematic overestimation across all dimensions. Students rated their critical evaluation competencies at or slightly above the midpoint of the scale (M range = 2.9–3.6 on a 5-point scale) yet showed failure-level performance on the corresponding tasks (M range = 18.3–38.5%), resulting in metacognitive gaps of 11.5 to 38.3 percentage points across all five dimensions.

Paired-samples t-tests confirmed that students significantly overestimated their abilities on every dimension (all ps < .001), with gaps ranging from 11.5 to 38.3 percentage points. Effect sizes were large across all dimensions (d = 1.11–2.25), indicating substantial discrepancies between perceived and actual competence. Students rated themselves most confident in argument identification (M = 3.6, 65.0%) despite this dimension yielding among the lowest actual performance (26.7%), producing the largest metacognitive gap of 38.3 percentage points. Even the smallest gap — multi-source synthesis at 11.5 percentage points — carried a large effect size (d = 1.11), underscoring the systematic and pervasive nature of metacognitive inaccuracy across the instrument.

Table 2
Metacognitive Accuracy: Self-Ratings Versus Actual Performance

Dimension	Self-Rating (M)	Self-Rating (%)	Actual (%)	Gap (pp)	t(29)	Cohen's d
Argument Identification	3.6	65.0	26.7	+38.3	12.34**	2.25
Evidence Evaluation	3.1	52.5	34.3	+18.2	7.21**	1.32
Assumption Recognition	2.9	47.5	18.3	+29.2	8.63**	1.58
Multi-Source Synthesis	3.0	50.0	38.5	+11.5	6.08**	1.11
Source Criticism	3.2	55.0	28.9	+26.1	9.17**	1.67

Note. Self-ratings were converted to percentages using the formula (rating − 1) / 4 × 100, mapping the scale endpoints (1 = 0%, 5 = 100%) for direct comparison with task accuracy scores. Gap = self-rating (%) minus actual performance (%); positive values indicate overestimation. Paired-samples t-tests tested whether mean gaps differed significantly from zero. pp = percentage points. $p < .001$ for all paired comparisons.

DISCUSSION

This study investigated whether Moroccan first-year CPGE students possess the critical evaluation competencies needed to engage safely and effectively with AI-generated academic content. Results revealed systematic evaluation failures across all dimensions (M accuracy = 18–39%), with particularly severe deficits in recognizing assumptions (18.3%) and evaluating evidence (34.3%). Additionally, students demonstrated profound metacognitive gaps, overestimating their competencies by 11–38 percentage points. These findings indicate that students lack both the analytical skills and the self-awareness necessary for AI literacy, creating conditions in which they will confidently accept AI outputs they cannot evaluate.

The AI Literacy Crisis: Unprepared Students Meet Generative AI

Our findings reveal more than critical reading deficiencies; they document an AI literacy crisis unfolding as generative AI proliferates in educational contexts, even as students lack the competencies to engage critically with AI-generated content. Students who cannot evaluate evidence quality (34% accuracy), detect unstated assumptions (18% accuracy), or verify source claims (29% accuracy) now encounter ChatGPT generating summaries for research papers, producing explanations of complex concepts, and creating study materials. The question is not hypothetical: what happens when these evaluation-deficient students encounter AI systems engineered to sound authoritative regardless of accuracy?

Large language models confidently generate false information—hallucinations in which AI asserts claims, cites studies, and presents statistics without uncertainty markers, even when the information is entirely fabricated (Ji et al., 2023). Evidence evaluation results are particularly troubling: only 34.3% of students achieved minimum proficiency (Table 1), and the remaining 65.7% accepted inadequately supported claims in our assessment materials, indicating they lack the skills to detect when evidence is weak, irrelevant, or fabricated. These students will incorporate AI hallucinations into their academic work, treating fabricated claims as factual because confident presentation overrides their limited evaluation capacities.

Assumption recognition showed the most severe deficit: only 1 student (3.3%) achieved minimum proficiency (Table 1), and the remaining 96.7% failed to identify unstated premises, treating explicitly stated claims as the entire argument. The assumption recognition finding is equally troubling in AI contexts.

AI models inherit biases from training data that predominantly reflect Western, Anglophone, Global North perspectives (Bender et al., 2021), leading them to make unstated assumptions about "normal" educational systems, "typical" economic development, and "appropriate" communication styles. If students cannot detect assumptions even in human-authored texts designed to make them recognizable, they certainly cannot identify when AI outputs embed culturally specific training biases. For Moroccan students navigating multilingual, postcolonial contexts, uncritical acceptance of AI-embedded assumptions undermines the capacity to engage using their own situated knowledge and cultural frameworks.

Students rated source criticism as their strongest evaluation skill (self-rating 3.2/5) yet achieved only 29% accuracy—a 26-percentage-point metacognitive gap. This disconnect is especially dangerous with AI-generated content. When ChatGPT cites sources, students will accept those citations at face value, failing to verify whether sources exist, whether they assert what ChatGPT claims, or whether they originate from credible outlets. The Dunning-Kruger effect (Kruger & Dunning, 1999) captures the compounding danger: students who cannot accurately assess their own critical evaluation abilities also cannot accurately judge AI output quality. When overconfident students accept overconfident AI outputs, and neither party signals its own limitations, the resulting conditions are more dangerous than poor performance alone.

Theoretical Implications

Our findings extend theoretical understanding of critical literacy in several ways. First, they demonstrate that critical reading comprises distinct competencies that develop somewhat independently: students showed relative strength in multi-source synthesis (38.5%) but performed poorly on assumption recognition (18.3%), supporting Grabe and Stoller's (2020) componential framework while challenging the assumption that competence in one dimension transfers to others. Second, the results reveal that metacognitive awareness is a critical yet underdeveloped component, consistent with Paul and Elder's (2006) emphasis on metacognitive monitoring as central to expert reading.

Third, the findings complicate discussions of cultural factors in critical thinking development. Atkinson (1997) questioned whether critical thinking, as conceptualized in Western traditions, transfers to non-Western contexts. However, our data suggest that pedagogical rather than cultural factors primarily drive the observed deficits: students did not reject critical evaluation as culturally inappropriate but simply lacked skills despite strong academic motivation. This aligns with Manalo and Sheppard (2016), who demonstrated that, with appropriate instruction, students from diverse cultural backgrounds develop critical literacy effectively. The issue appears to be the instructional approach—emphasizing memorization and reproduction over analysis and evaluation—rather than cultural resistance.

Practical Implications: A Diagnostic-First Approach to AI Integration

These findings carry urgent practical implications for educational institutions considering or implementing AI integration. We propose a three-phase diagnostic-first approach.

Phase 1: Diagnostic Assessment

Before introducing AI tools, institutions should systematically assess students' critical evaluation competencies. This diagnostic phase identifies specific gaps—such as difficulty evaluating evidence, inability to recognize assumptions, and lack of source verification skills. Without diagnostic data, institutions risk wasting instructional time on skills students already possess while neglecting areas of genuine need. Our CRDA offers one approach, though the principle matters more than any specific tool: measure actual competencies rather than assuming students possess skills because they completed prior education.

Phase 2: Targeted Instruction

Diagnostic results should inform explicit, sustained instruction that addresses identified gaps. For assumption recognition (96.7% failure rate; Table 1), instruction should include explicit teaching of common assumption types, practice identifying unstated premises in progressively more complex texts, and metacognitive reflection on when assumptions go unnoticed. For evidence evaluation (65.7% accepting inadequate support; Table 1), instruction should emphasize distinguishing among evidence types and recognizing when confident presentation masks evidentiary weakness. Critically, instruction must address metacognitive dimensions alongside analytical skills: students need not only the capacity to evaluate evidence but also a realistic awareness of their own evaluation limitations.

Phase 3: AI Integration with Explicit Evaluation Frameworks

Only after students demonstrate baseline competencies should institutions integrate AI tools, and this integration must include explicit frameworks for evaluating AI-generated content. Students need to understand that AI lacks semantic comprehension, cannot distinguish verified facts from plausible fabrications, and inherits training biases. They need specific strategies: always verify AI citations, cross-check AI claims against reliable sources, recognize that fluent writing does not ensure accurate content, and treat AI as a starting point that requires critical assessment. Technology-first approaches that assume students will develop evaluation skills through AI use itself are directly challenged by our data.

Limitations

Several limitations warrant acknowledgment. First, the sample comprised 30 students from a single CPGE program, limiting generalizability. While CPGE students' academic selectivity makes the findings concerning—if the top 10% struggle this severely, broader populations likely face even greater challenges—conclusions should be tested across multiple programs and educational levels. Second, our assessment instrument mirrors AI characteristics using researcher-designed content rather than authentic AI outputs, raising questions about ecological validity; however, if students cannot evaluate content designed to be evaluable, authentic AI outputs would likely decrease rather than improve performance. Third, our cross-sectional design cannot determine how competencies might develop with CPGE instruction. Longitudinal research tracking students across the program would illuminate whether explicit critical reading instruction addresses the identified deficits.

CONCLUSION

We began this study with a question made urgent by the rapid proliferation of generative AI: Are Moroccan CPGE first-year students prepared to critically engage with AI-generated academic content? Our data provide unambiguous answers: they are not. Students demonstrated systematic evaluation failures across all five critical dimensions we assessed, with mean accuracy rates of only 18–39%. Most concerning, students cannot accurately judge their own evaluation abilities, overestimating their competencies by 11–38 percentage points across all dimensions. This convergence of poor performance and profound metacognitive gaps creates conditions in which students will confidently accept AI outputs they cannot evaluate.

Three patterns emerged clearly. First, assumption recognition proved alarming: 96.7% of students failed to identify unstated premises, even in texts designed to make assumptions clear. Second, evidence evaluation showed severe deficiencies: only 34.3% of students could identify inadequately supported claims, revealing a widespread inability to distinguish well-founded assertions from confident-sounding fabrications. Third, metacognitive gaps proved as severe as performance deficits: students consistently rated themselves as moderately competent while performing at failure levels.

These findings extend beyond Morocco. They document a global educational challenge emerging wherever generative AI meets students who lack a foundation in critical evaluation. If Morocco's highest academic achievers—the top 10% of national graduates who have succeeded through intensely competitive selection—cannot critically evaluate AI-generated content, what about broader student populations? The implications are alarming: rushing AI integration without diagnostic assessment risks amplifying rather than addressing educational inequalities.

Educational institutions face a fundamental choice: continue current technology-first AI integration, assuming students possess evaluation competencies, or adopt diagnostic-first AI integration that assesses students' actual competencies before introducing AI tools. Our Critical Reading Diagnostic Assessment shows that such instruments can be developed, validated, and administered efficiently. The broader principle matters more than any specific tool: understand students' actual competencies before integrating technologies that assume those competencies. Diagnosis first, technology second—this is what responsible AI integration in education requires, and our students' futures depend on getting this sequence right.

Author Contributions: Conceptualization, Z.E.H.; methodology, Z.E.H.; validation, Z.E.H.; formal analysis, Z.E.H.; investigation, Z.E.H.; data curation, Z.E.H.; writing—original draft preparation, Z.E.H.; writing—review and editing, N.H.; supervision, N.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the ethical principles of the Declaration of Helsinki and received institutional approval prior to data collection. Participation was voluntary, participants were assured that their performance would not affect their grades, and all data were anonymized before analysis. No formal approval number was assigned.

Informed Consent Statement: Informed consent was obtained from all participants involved in the study.

Acknowledgments: The authors thank the first-year CPGE students who participated in this study, the three faculty members who reviewed the CRDA for content validity, and the second rater who contributed to the inter-rater reliability analysis. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

REFERENCES

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Alwadai, M. (2014). *Islamic teachers' perceptions of improving critical thinking skills in Saudi Arabian elementary schools* [Doctoral dissertation, Southern Illinois University Carbondale]. OpenSIUC. http://opensiuc.lib.siu.edu/dissertations/882

Amrous, N., & Nejmaoui, N. (2016). A developmental approach to the use of critical thinking skills in writing: The case of Moroccan EFL university students. In *Arab World English Journal: ASELS Annual Conference Proceedings* (pp. 142–156). AWEJ.

Atkinson, D. (1997). A critical approach to critical thinking in TESOL. *TESOL Quarterly, 31*(1), 71–94. https://doi.org/10.2307/3587975

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). ACM. https://doi.org/10.1145/3442188.3445922

Buselic, V. (2023). Teaching information literacy and critical thinking skills in ChatGPT time. In *2023 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA)* (pp. 14–20). IEEE. https://doi.org/10.1109/CoNTESA61248.2023.10384921

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Conseil Supérieur de l'Éducation, de la Formation et de la Recherche Scientifique. (2015). *Vision stratégique de la réforme 2015–2030: Pour une école de l'équité, de la qualité et de la promotion*. Royaume du Maroc. https://www.csefrs.ma/wp-content/uploads/2017/09/Vision_VF_Fr.pdf

Darvin, R. (2025). The need for critical digital literacies in generative AI-mediated L2 writing. *Journal of Second Language Writing, 67*, Article 101186. https://doi.org/10.1016/j.jslw.2025.101186

Grabe, W., & Stoller, F. L. (2020). *Teaching and researching reading* (3rd ed.). Routledge.

Hellalet, N. (2021). Critical thinking in Moroccan university students' argumentative writing. *Journal of Applied Language and Culture Studies*, *4*, 179–194.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), Article 248. https://doi.org/10.1145/3571730

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Kurland, D. J. (2000). *How the language really works: The fundamentals of critical reading and effective writing*. CriticalReading.com. http://www.criticalreading.com/critical_reading.htm

Liu, F., & Stapleton, P. (2014). Counterargumentation and the cultivation of critical thinking in argumentative writing: Investigating washback from a high-stakes test. *System, 45*, 117–128. https://doi.org/10.1016/j.system.2014.05.005

Manalo, E., & Sheppard, C. (2016). How might language affect critical thinking performance? *Thinking Skills and Creativity, 21*, 41–49. https://doi.org/10.1016/j.tsc.2016.05.005

Melisa, R., Yenti, L., Nurwahyuningsih, & Sari, D. P. (2025). Critical thinking in the age of AI: A systematic review. *Educational Process: International Journal, 14*, Article 31. https://doi.org/10.22521/edupij.2025.14.31

Paul, R., & Elder, L. (2006). *Critical thinking: Tools for taking charge of your learning and your life* (2nd ed.). Pearson Prentice Hall.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Lawrence Erlbaum Associates.

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems, 3*, 121–154. https://doi.org/10.1016/j.iotcps.2023.04.003

Sardi, J., Darmansyah, C., Candra, O., Yuliana, D. F., Habibullah, Yanto, D. T. P., & Eliza, F. (2025). How generative AI influences students' self-regulated learning and critical thinking skills? A systematic review. *International Journal of Engineering Pedagogy, 15*(1), 94–108. https://doi.org/10.3991/ijep.v15i1.53379

Thongsan, S., & Anderson, P. (2025). How AI supports critical reading in EFL classrooms. *LEARN Journal: Language Education and Acquisition Research Network, 18*(2), 795–820. https://files.eric.ed.gov/fulltext/EJ1479855.pdf

Wang, C., & Wang, Z. (2025). Investigating L2 writers' critical AI literacy in AI-assisted writing: An APSE model. *Journal of Second Language Writing, 67*, Article 101187. https://doi.org/10.1016/j.jslw.2025.101187

Wang, Y., Pan, Y., Yan, M., Su, Z., & Luan, T. H. (2023). A survey on ChatGPT: AI-generated contents, challenges, and solutions. *IEEE Open Journal of the Computer Society, 4*, 280–302. https://doi.org/10.1109/OJCS.2023.3300321

Xiao, F., Zhu, S., & Xin, W. (2025). Exploring the landscape of generative AI (ChatGPT)-powered writing instruction in English as a foreign language education: A scoping review. *ECNU Review of Education*. Advance online publication. https://doi.org/10.1177/20965311241310881