مجلة اللسان الدولية للدراسات اللغوية والأدبية
AL-LISAN INTERNATIONAL JOURNAL FOR LINGUISTIC & LITERARY STUDIES

Vol 10, No 22,(SEPTEMBER 2026) (AIJLLS) المجلد 10، العدد 22 (سبتمبر 2026)

من موضوعات العدد

- التناوب الدلالي بين أبنية المصدر: دراسة صرفية دلالية
- الصراع الهوياتي بين الشخصيات في رواية: لا احبّ الشمس في باريس
- صورة الفرس في شعر امرئ القيس وعنترة بن شداد: دراسة في الوسائط اللغوية والأنساق المضمرة
- دور الألعاب الإلكترونية التعليمية في تنمية الإنتاج الشفهي لدى متعلمي اللغة العربية الناطقين بغيرها

تصدرها
PUBLISHED BY
كلية اللغات – جامعة المدينة العالمية
FACULTY OF LANGUAGES
AL-MADINAH INTERNATIONAL UNIVERSITY

جامعة المدينة العالمية
Al-Madinah International University

AIJLLS |SEPTEMBER 2026 | VOL 10 ISSUE 22
مجلة اللسان الدولية / العدد الثاني والعشرون (سبتمبر 2026م/ 1448 هـ)
ISSN 2600-7398 الترقيم الدولي

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

An Investigation into the Role of Large Language Models (LLMs) in Cultivating and Fostering Some of the EFL 21st Century Skills among Al Madinah International University Students

Gellan Gamal Atwa[1*]

[1] Lecturer, Al-Madinah International University

[*] gellan.abdel@mediu.my

Abstract

This research investigates the impact of Large Language Models (LLMs) on the development of 21st-century skills, particularly critical thinking and creativity, among English as a Foreign Language (EFL) students at Al Madinah International University. Given the global importance of English proficiency and the educational potential of LLMs like ChatGPT and Gemini, this study fills a gap in existing research by directly investigating the empirical effect of LLMs on the four key skills (critical thinking, creativity, communication, and collaboration) beyond basic language tasks. The study employed a quasi-experimental design, with an experimental group (n = 17) utilising LLMs and a control group (n = 17) employing traditional teaching methods. A comprehensive 21st-century skills test, validated for expert review, internal consistency (Cronbach's Alpha for critical thinking: 0.704, creativity: 0.906, and overall: 0.885), and discriminant validity (significant differences between high and low scores at $\alpha \leq 0.01$), was administered as pre- and post-tests. Test-retest reliability further confirmed the results (critical thinking: 0.767, creativity: 0.713, overall: 0.894). Statistical analysis, using the Mann-Whitney U Test and Wilcoxon Signed Ranks Test, showed significant differences ($\alpha \leq 0.01$) in favour of the experimental group on post-test scores for critical thinking and creativity. The substantial effect sizes indicate a strong positive impact of integrating LLMs on these skills. These findings suggest that LLMs can effectively enhance critical thinking and creativity in EFL education, offering practical insights for educators and policymakers seeking to incorporate advanced technological solutions into language-learning curricula.

ملخص البحث:

تستقصي هذه الدراسة تأثير نماذج اللغات الكبيرة (LLMs) على تطوير مهارات القرن الحادي والعشرين، لا سيما التفكير النقدي والإبداع، لدى طلاب اللغة الإنجليزية لغة أجنبية (EFL) في الجامعة الإسلامية العالمية بالمدينة المنورة. ونظراً لأهمية الكفاءة في اللغة الإنجليزية عالمياً والإمكانات التعليمية لنماذج اللغات الكبيرة مثل ChatGPT وجيمني (Gemini)، تسد هذه الدراسة فجوة في البحوث القائمة من خلال التحقيق المباشر في الأثر التجريبي لنماذج اللغات الكبيرة على المهارات الأربع الرئيسية (التفكير النقدي، الإبداع، التواصل، التعاون) متجاوزةً المهام اللغوية الأساسية. استخدمت الدراسة تصميماً شبه تجريبي، شمل مجموعة تجريبية (العدد = 17) تستخدم نماذج اللغات

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

الكبيرة ومجموعة ضابطة (العدد = 17) تستخدم طرق التدريس التقليدية. وتم تطبيق اختبار شامل لمهارات القرن الحادي والعشرين –الذي تم التحقق من صدقه عبر مراجعة الخبراء، والاتساق الداخلي (معامل ألفا كرونباخ للتفكير النقدي: 0.704، والإبداع: 0.906، والإجمالي: 0.885)، والصدق التمايزي (فروق ذات دلالة بين الدرجات العالية والمنخفضة عند مستوى دلالة $\alpha \le 0.01$)– كاختبارين قبلي وبعدي. كما أكدت موثوقية اختبار–إعادة الاختبار النتائج (التفكير النقدي: 0.767، الإبداع: 0.713، الإجمالي: 0.894). أظهر التحليل الإحصائي، باستخدام اختبار مان–ويتني (Mann–Whitney U Test) واختبار ويلكوكسون للرتب الموقعة (Wilcoxon Signed Ranks Test)، فروقاً ذات دلالة إحصائية ($\alpha \le 0.01$) لصالح المجموعة التجريبية في درجات الاختبار البعدي للتفكير النقدي والإبداع. وتُشير أحجام الأثر الكبيرة إلى تأثير إيجابي قوي لدمج نماذج اللغات الكبيرة على هذه المهارات. وتقترح هذه النتائج أن نماذج اللغات الكبيرة يمكنها تعزيز التفكير النقدي والإبداع بفعالية في تعليم اللغة الإنجليزية لغة أجنبية، مما يوفر رؤى عملية للمعلمين وصنّاع القرار الساعين إلى دمج الحلول التكنولوجية المتقدمة في مناهج تعلم اللغات.

Introduction

The global significance of the English language is underscored by its pervasive role in commerce, pedagogy, scientific inquiry, and international discourse. This ubiquity compels both individuals and states to prioritize English proficiency as a fundamental vehicle for fostering academic, vocational, and interpersonal advancement. Notwithstanding the multifaceted hurdles associated with English as a Foreign Language (EFL) acquisition—such as catering to diverse learner profiles and the imperative of robust pedagogical motivation—the momentum of globalization ensures that high-level English competency remains a requisite skill for meaningful global participation.

EFL proficiency has emerged as a cornerstone of modern international infrastructure, functioning as a *lingua franca* across scientific, technological, and educational domains. Mastery of the language facilitates intercultural exchange and unlocks significant career and academic trajectories, frequently serving as a mandatory criterion for higher education and professional recruitment (San Jose, Guitche, & Madrigal, 2024). Consequently, pedagogical standards such as the CEFR have been widely adopted across diverse educational landscapes. Nevertheless, the pursuit of linguistic competence is often impeded by variables including performance anxiety, a dearth of authentic communicative opportunities, and the difficulty of sustaining long-term learner motivation (Nguyen, 2024). A spectrum of factors, ranging from demographic characteristics to institutional support and instructor availability, further modulates proficiency outcomes. While contemporary advancements in digital technology and artificial intelligence have begun to augment the language-learning experience, the development of English skills remains a paramount educational objective, increasingly viewed as essential for both global citizenship and individual development (Dan & Huang, 2024).

Large language models (LLMs), such as Gemini and ChatGPT, have demonstrated substantial progress in producing and interpreting

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

human-like textual data across an expansive array of disciplines. Their capacity to provide granular explanations and resolve complex inquiries has facilitated their adoption within specialized sectors including medicine, jurisprudence, and commerce. Within academia, these models are increasingly leveraged as auxiliary pedagogical instruments, demonstrating notable efficacy in professional assessment environments and test preparation. The notable performance of tools like ChatGPT on standardized medical examinations illustrates their utility for both academic aspirants and instructors. Although LLMs represent a paradigm shift in information accessibility and practical application, their deployment requires cautious calibration. The presence of occasional inaccuracies and systemic inconsistencies necessitates stringent human oversight to guarantee reliability and accuracy. Consequently, while LLMs possess the potential to redefine learning architectures, a dependable and strategic integration remains essential (Koga, 2023).

The integration of technology has become a hallmark of contemporary EFL instruction, employing digital resources to refine language acquisition, bolster student engagement, and cultivate 21st-century competencies. Although learners frequently report that digital integration promotes motivation and facilitates personalized, collaborative learning environments, its application is not devoid of challenges. Primary impediments include disparate levels of digital literacy among learners, inequitable resource access, and the critical requirement for ongoing professional development for educators. The COVID-19 pandemic acted as an accelerant, forcing an industry-wide transition toward online and hybrid instructional models, thereby exposing both the affordances and the limitations of this digital paradigm. The efficacy of these technological tools is fundamentally contingent upon institutional support, teacher training, and prevailing attitudes toward innovation. Currently, the maturation of LLMs and broader AI systems heralds a new era, offering the potential to further revolutionize EFL instruction through sophisticated and interactive language scaffolding (Tawafak, Al-Obaydi, Klimova & Pikhart, 2023; Nurhidayat, Mujiyanto, Yuliasri, & Hartono, 2024).

Literature Review

For students acquiring English as a Second Language, proficiency in critical thinking, creativity, communication, and collaboration constitutes the cornerstone of 21st-century competency. There is a broad consensus that these attributes are indispensable for future success, particularly as the global landscape grows increasingly complex and technologically integrated. Critical thinking is fundamentally characterized by the ability to objectively appraise information, detect implicit biases, and derive logical inferences. Simultaneously, creativity is manifested through innovative problem-solving and the generation of original, imaginative concepts—both of which are increasingly prioritized in modern pedagogical frameworks. While communicative competence encompasses a diverse array of skills, such as precise written discourse, effective nonverbal cues, and articulate oral expression, collaborative teamwork emphasizes the collective synergy required to achieve shared objectives. These interrelated proficiencies are not merely supplementary; they are essential for navigating and surmounting the multifaceted challenges of a diverse global society.

Critical thinking serves as a multifaceted process, vital for EFL learners as they navigate and analyze complex texts to

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

discern authorial intent and profound thematic concepts (Negoescu, 2023). The subsequent phase involves evaluation, wherein students scrutinize source credibility and argumentative coherence while remaining vigilant against potential biases (Peng, 2023; Wangmanee & Kaowiwattanakul, 2024). Synthesis requires the integration of disparate knowledge bases to construct unique, persuasive arguments (Darwin et al., 2024). Furthermore, active inquiry through the formulation of probing questions serves to deepen engagement with instructional material (Azzalini, 2023; Shalikiani-Skhireli, 2021). Ultimately, this cognitive process empowers students to apply logical rigor to reach valid conclusions, grasp broader implications, and articulate robust arguments in English (Azzalini, 2023).

Critical thinking is a key 21st-century attribute for informed decision-making grounded in research and rigorous evaluation (Todorovska, 2024). EFL pedagogy transcends rote memorization, necessitating a meticulous examination, assessment, and assimilation of linguistic content. By fostering a more profound engagement with the language, this approach enhances proficiency in speaking, listening, writing, and reading, thereby encouraging reflective learning (Lan, 2024; Mzeil, 2024; Todorovska, 2024). The integration of critical thinking into EFL curricula is therefore of paramount importance (Wangmanee & Kaowiwattanakul, 2024). Moreover, it is central to academic achievement, equipping students with the capacity to resolve complex problems, exercise logical judgment, and interact effectively within cross-cultural contexts. By championing process-oriented literacy and validating a plurality of perspectives, critical thinking also catalyzes diversity (Mehta, 2024).

To cultivate critical thinking within EFL environments, educators may deploy a range of instructional strategies. For instance, reading comprehension can be significantly bolstered through direct pedagogical intervention. According to Ferdoush and Jahan (2024), this approach involves scaffolding students' use of specific techniques, such as Critical Discourse Analysis, alongside the use of inquiry-based questioning. Utilizing authentic resources—such as journalism and literary works—exposes students to varied viewpoints and real-world linguistic applications. Furthermore, participation in structured debates and communicative exercises strengthens their capacity to synthesize, analyze, and defend arguments (Darwin et al., 2024). Project-Based Learning (PBL) and reflective journaling are further instrumental strategies that enhance both language proficiency and critical thinking by facilitating introspection and collaborative problem-solving (Eto, 2022). While there remains significant scope for improvement—particularly regarding curriculum reform and continuous professional development—AI can serve to complement, rather than displace, human critical thinking. Nevertheless, persisting issues such as cultural variances and educator preparedness warrant continued attention.

Creativity within the EFL paradigm encompasses an expansive spectrum of activities, including divergent thinking, the

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

inventive application of language, and the conceptualization of novel ideas. Now firmly established as an essential 21st-century skill, creative methodologies are increasingly utilized to elevate student engagement, nurture critical inquiry, and accelerate language proficiency (Rawlings et al., 2025). In this context, creativity is defined as the capacity to generate original ideas, experiment with linguistic nuances, and employ divergent thinking to approach problems from varied perspectives. Such creative engagement in EFL settings typically incorporates brainstorming, narrative construction, gaming, and digital storytelling—all of which encourage students to expand their lexicon, articulate diverse viewpoints, and develop innovative communicative strategies (Yeh & Chen, 2024).

Divergent thinking, the innovative application of language, and the generation of original concepts constitute the tripartite foundation of creativity within EFL pedagogy. The primary objective of divergent thinking is to produce a breadth of ideas, a process fortified by activities that incentivize unconventional cognitive patterns. This serves as a necessary complement to convergent thinking, which involves the critical appraisal and selection of the most viable solutions. Furthermore, "novel language use" involves the creative repurposing of vocabulary, while linguistic play serves as a mechanism to unlock creative expression. The synergy between these processes fosters an environment conducive to innovative problem-solving. Ultimately, originality—defined as the development of novel ideas—remains a central component of EFL creativity, particularly when applied to linguistic articulation and complex problem-solving tasks (Alfares, 2024; Tin, 2024).

An instructor's autonomy, passion, and creative self-efficacy are foundational to cultivating such environments, as these psychological variables directly dictate their capacity to design and execute pedagogical innovation. Digital technologies, such as virtual reality and AI-integrated tools, can significantly amplify motivation and cognitive flexibility by providing immersive, interactive, and collaborative platforms for language practice. However, achieving a sustainable equilibrium between human instruction and AI assistance requires careful strategic oversight. Innovation in the EFL sector is frequently hindered by structural constraints, including rigid curricula, examination-oriented systems, time limitations, and digital inequality. Meaningful pedagogical creativity necessitates the application of metacognitive techniques, the cultivation of supportive learning environments, and a willingness to experiment with and refine instructional approaches. Conclusively, fostering creativity within EFL not only accelerates language proficiency but also equips students with the critical thinking competencies required to navigate an increasingly volatile global environment (Behzadpoor, Fathi & Greenier, 2023; Wang, 2024; Alzubi, Alyami, & Nazim, 2025).

In the design of EFL curricula, educators should prioritize the integration of critical thinking, creativity, and innovative linguistic application. Beyond merely enhancing their conceptual understanding of creativity, teacher preparation programs must provide

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

practitioners with actionable, high-quality resources. Incorporating critical thinking into EFL instruction is imperative for advancing both language acquisition and complex problem-solving capabilities. By systematically implementing these pedagogical strategies, EFL practitioners can curate stimulating learning environments that nurture student creativity and elevate overall language proficiency, resulting in a more enriched educational experience (Jabeen & Sarifa, 2024; Li & Wei, 2025).

Robust assessment methodologies are essential for evaluating collaborative and communicative competencies, both of which are critical in academic and professional spheres. Key evaluative instruments include observation checklists, peer-review forms, and the systematic analysis of communication records (Madi et al., 2024). The efficacy of peer assessment can be significantly improved through the utilization of structured rubrics and comprehensive training; this enables students to accurately appraise peer contributions while gaining deeper insight into group dynamics and individual accountability (Madi et al., 2024). Within group activities, observation checklists provide a systematic framework for monitoring specific skills, including interpersonal communication and task management (Arthars et al., 2024; Demetroulis et al., 2024). Furthermore, the analysis of communication logs from collaborative platforms offers instructors a granular record of interactions, elucidating patterns of engagement and the qualitative nature of student discourse (Khalil, 2015).

Large Language Models (LLMs) may be successfully integrated into educational settings when grounded in rigorous theoretical frameworks. Constructivism, connectivism, and socio-cultural theory offer multifaceted perspectives on the learning process and the strategic application of LLMs (Nithideechaiwarachok et al., 2024; Vendri et al., 2024; Karki & Karki, 2024). By synthesizing constructivist and connectivist paradigms, researchers can establish a robust foundation for understanding how LLMs catalyze learning in the digital age (Hendricks, 2019; Goldie, 2016).

As noted by Hendricks (2019) and Jeny (2024), connectivism emphasizes the critical importance of evaluating networks, interpersonal relationships, and collaborative effort within digital learning environments. This theory advocates for the cultivation and utilization of distributed networks to enhance learning, shifting the focus away from traditional, didactic information-delivery models toward a more dynamic, interconnected pedagogical approach (Putu, 2024).

Connectivism posits that learning is an emergent process occurring through the navigation of interconnected networks, where knowledge is distributed rather than centralized. The capacity to continuously learn is paramount, as knowledge resides within these networks; articulating these connections is vital for fostering lifelong learning (Hendricks, 2019). Within this framework, Large Language Models (LLMs) are instrumental, serving as catalysts that grant learners access to expansive information reservoirs and facilitate connections to pertinent resources. LLMs enhance educational delivery through several mechanisms:

Networked Learning: Students can interact with a diverse array of knowledge sources

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

and domain experts, thereby broadening their intellectual horizons (Goldie, 2016).

Digital-Age Compatibility: As active participation in digital ecosystems is essential for data synthesis and meaning-making, LLMs are naturally aligned with contemporary digital requirements (Kim, 2022).

Resource Development: These models assist in defining learning objectives, generating visual aids, and customizing content to foster immersive learning experiences (Ballard et al., 2025).

Entrepreneurial Skill Acquisition: By providing access to real-time information and simulating professional interactions, LLMs assist individuals in skill development and the cultivation of an entrepreneurial mindset (Thanasi-Boçe & Hoxha, 2024).

Conversely, constructivism defines learning as an active, recursive process of meaning-making. From this perspective, students acquire knowledge by engaging with their environment, peers, and lived experiences (Vendri et al., 2024). This pedagogical philosophy prioritizes scaffolding—whereby learners are guided by more experienced counterparts (Karki & Karki, 2024)—as well as the significance of social interaction in knowledge construction (Vendri et al., 2024) and active discovery through experimentation (Gouloudis, 2020). LLMs align with constructivist frameworks by providing opportunities for active participation (Vendri et al., 2024), personalized experiences tailored to individual learner needs (Thanasi-Boçe & Hoxha, 2024), and allowing educators to function as facilitators who guide students in deepening their conceptual understanding (Chowdhury et al., 2020).

Key components of constructivist learning environments include student collaboration, shared control, personal relevance, an openness to uncertainty, and the promotion of critical discourse. These elements positively influence both intrinsic motivation and extrinsic achievement, ultimately shaping cognitive outcomes such as elaboration, critical thinking, and metacognitive self-regulation (Do et al., 2023). Constructivist principles are also applicable to non-formal education, highlighting the interplay between the learner and their environment (Vendri et al., 2024). However, learners accustomed to traditional didactic methods often require targeted support to embrace the challenges inherent in self-directed learning (Chowdhury et al., 2024).

LLMs provide a suite of features that are advantageous to both educators and learners, particularly within language acquisition contexts. These models can condense complex texts, thereby improving comprehension and retention, and excel at generating reading lists, exercises, and multi-level content. Their role-playing capabilities simulate realistic dialogue, while question-and-answer functions enable students to verify understanding and resolve uncertainties in real-time. Although LLMs demonstrate sophisticated command of grammar and syntax, challenges remain regarding their application of practical knowledge to complex, real-world scenarios.

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

Current research is actively addressing limitations in knowledge transfer and contextual nuance. These models are highly scalable and customizable for diverse languages and multimedia formats, encompassing text, visual media, and audio (Mukta et al., 2024; Mahowald, 2024; Patil & Gudivada, 2024).

Linders and Tomczak (2025), Zhou (2025), and Sinh (2025) assert that LLMs possess the potential to revolutionize EFL instruction. They can generate a breadth of educational resources, from pedagogical explanations to personalized textbooks derived from existing presentations. A core capability of these models is text summarization; by employing techniques such as QA-prompting, they significantly enhance the quality of these condensations. In essence, LLMs serve as responsive agents for inquiries regarding linguistics, grammar, and vocabulary. By integrating natural language understanding with knowledge graphs, they improve their ability to address complex inquiries. Furthermore, LLMs facilitate a low-anxiety environment for speaking and listening practice through accurate conversation exercises and role-play. By adapting to individual learning styles through customized activities and adaptive tools, they markedly improve the efficacy of online language learning experiences.

The integration of Large Language Models (LLMs) is increasingly critical for the advancement of EFL pedagogy. Recent empirical investigations underscore their substantial impact on writing proficiency, grammatical accuracy, communicative practice, and lexical acquisition. In structured learning environments, LLMs bolster student motivation, facilitate the development of writing skills, and incentivize self-directed learning. Although these models necessitate careful calibration for nuanced evaluative tasks, they demonstrate high reliability in grading essays and providing immediate, formative feedback, thereby supporting personalized learning trajectories. Furthermore, they enhance recall and mitigate orthographic and lexical errors, with learners frequently reporting high levels of satisfaction. While feedback can occasionally be concise, LLMs cultivate low-anxiety environments for speaking practice. Ultimately, the perceived utility of these tools significantly influences learner motivation and acceptance. Instructors continue to emphasize the necessity of deliberate integration and ongoing collaboration with educators to effectively leverage these benefits while mitigating the inherent risks associated with over-reliance on LLM technologies.

LLMs provide significant support for academic writing. By offering structural suggestions and enhancing prose quality, these models assist EFL students in the iterative drafting process. LLMs serve as effective resources for scaffolding, helping students generate preliminary versions that they can subsequently refine. It is imperative, however, to acknowledge the phenomenon of "hallucination"—the tendency of LLMs to generate technically plausible but factually erroneous information. Consequently, students must critically evaluate AI-generated outputs rather than accepting them at face value (Kumar, 2024). Furthermore,

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

grammar correction remains one of the most functional applications of LLMs for EFL learners. These models identify and rectify syntactical errors while offering explanations that foster a deeper understanding of English grammar. The efficacy of this feedback is contingent upon the accuracy of the LLM and its capacity to provide insightful, metalinguistic commentary. Research into recasts and error-corrective feedback has shown significant promise in enhancing the metalinguistic awareness and grammatical performance of EFL learners (Barzang, 2019).

EFL students may also leverage LLMs to advance their oral proficiency. Through interactive dialogue, learners receive real-time feedback on grammar, pronunciation, and fluency. Mobile-Assisted Language Learning (MALL) programs that incorporate LLMs have demonstrated considerable potential for enhancing speaking abilities by providing interactive exercises and tailored feedback, thereby increasing engagement (Hwang et al., 2024). Additionally, LLMs facilitate lexical expansion by defining new terms, providing contextual usage examples, and designing targeted assessments to solidify vocabulary retention. These models further enable individualized vocabulary instruction, catering to the specific requirements and preferences of the learner (Ding et al., 2025). While LLMs enhance EFL learning through personalized feedback and resource diversity, it is essential to systematically address the associated ethical challenges.

Despite the pedagogical advantages of LLMs, moral concerns remain salient (Haltaufderheide & Ranisch, 2024). A primary concern is the propensity for LLMs to reinforce biases inherent in their training datasets (Lin et al., 2024; Kumar, 2024). To promote equity and reduce systemic bias, it is recommended that LLMs be trained on more diverse and representative datasets (Lin et al., 2024). These ethical challenges are compounded by the potential for academic dishonesty. The use of LLMs to generate homework or essays threatens the integrity of the educational process. To mitigate this risk, educators must develop robust methods for identifying academic misconduct while simultaneously fostering a culture that encourages the ethical and appropriate application of AI (Chiang et al., 2022).

The extent to which the "four Cs"—critical thinking, creativity, communication, and collaboration—are impacted by the proliferation of LLMs in education remains under-researched. Current literature largely prioritizes the impact of LLMs on programming and code quality, effectively neglecting their role in developing these critical 21st-century competencies. This underscores an urgent need for targeted empirical research to investigate how LLMs can support these skills beyond basic language tasks. The present study aims to bridge this research gap, providing educators and policymakers with insights into LLM integration with a specific focus on the 4 Cs (Alias, 2024).

Drawbacks of LLMs, such as the generation of false information and significant computational energy requirements, present ongoing challenges to their responsible implementation (Kumar, 2024; Abdali et al.,

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

2024). Notwithstanding these limitations, LLMs possess the capacity to transform language instruction by enhancing pedagogical efficacy and providing deeply individualized learning opportunities (Zeng et al., 2024).

In summary, direct empirical research regarding the specific effects of LLMs on the development of the 4 Cs remains limited, extending well beyond rudimentary language practice. Although existing research addresses adjacent topics, this gap remains largely unexplored. For instance, studies on online professional development underscore the significance of technological proficiency for skill enhancement but often omit the 4 Cs (Meyer et al., 2023). Similarly, while research on governance in open-source software provides insights into the "Collaboration" element of the 4 Cs, it fails to investigate how LLMs directly influence teamwork and communicative dynamics (Li et al., 2021). Therefore, there is a clear imperative to focus explicitly on the intersection of LLM integration and the 4 Cs to advance both pedagogical theory and practice.

The research questions

1. **What is the impact of using LLMs on EFL students' critical thinking skills at Al Madinah International University?**
2. **To what extent do LLMs have an effect on EFL students' creativity skills at Al Madinah International University?**

Methods

A pilot sample of 15 male and female students from Al Madinah International University was given the test to ascertain the following:

a) Calculating the Validity of the Test:

The validity of the test was calculated using the following methods:

- **Expert Validity Method** Expert validity was used to ascertain the test's validity by presenting the test to a group of experts to obtain their opinions regarding:
    - Sufficiency of instructions provided to students to answer the test correctly.
    - Scientific and linguistic validity of the items.
    - Suitability of the items for the sample students.
    - Suitability of each question for the skill it was designed to measure.
    - Each question achieves its intended objective.

The experts agreed on the validity, suitability, and soundness of the test items.
The researcher used SPSS V.18 software to calculate the internal consistency validity of the test by computing:
a) Internal consistency between the score of each sub-skill and the total score of the primary skill to which that sub-skill belongs.
b) Internal consistency between the score of each primary skill and the total score of the dimension to which that primary skill belongs.
c) Internal consistency between the score of each dimension and the total score of the entire test.

Internal Consistency Between Each Sub-Skill Score and the Total Score of Its Main Skill

The internal consistency of the sub-skills was calculated by computing **Pearson correlation coefficients** between students' scores on each sub-skill and the total score of the primary skill to

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

illustrates the validity coefficients of the sub-skills for the 21st-century skills test:

Table (1)
Correlation Coefficient Between Each Sub-Skill Score and the Total Score of the Main Skill to Which It Belongs for the 21st-Century Skills Test (N = 20)

No	Main Skills		Sub-Skills	Correlation
1	Critical Thinking	Analysing and Evaluating Information	Sub-Skill 1	0.629**
2			Sub-Skill 2	0.734**
3			Sub-Skill 3	0.764**
4		Problem-Solving and Decision-Making	Sub-Skill 4	0.800**
5			Sub-Skill 5	0.678**
6	Creativity	Generating New Ideas and Solutions	Sub-Skill 6	0.909**
7			Sub-Skill 7	0.921**
8		Adapting and Iterating	Sub-Skill 8	0.934**
9			Sub-Skill 9	0.919**

**. Correlation is significant at the 0.01 level

Internal Consistency Between Each Main Skill Score and the Total Score of its Dimension

The internal consistency of the main skills was calculated by computing **Pearson correlation coefficients** between students' scores on each primary skill and the total score of the dimension to which that primary skill belongs. The table below presents the validity coefficients for the main skills of the 21st-century skills test:

Table (2)
Correlation Coefficient Between Each Main Skill Score and the Total Score of the Dimension to Which It Belongs for the 21st-Century Skills Test (N = 20)

No	Main Skills	Correlation

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

1	**Critical Thinking**	**Analysing and Evaluating Information**	0.960**
2		**Problem-Solving and Decision-Making**	0.907**
3	**Creativity**	**Generating New Ideas and Solutions**	0.961**
4		**Adapting and Iterating**	0.959**

**. Correlation is significant at the 0.01 level
All correlation coefficients are significant at the 0.01 level, as shown in the preceding table, demonstrating the internal consistency validity of the test's primary competencies.

The internal consistency for the test's dimensions was calculated by computing Pearson Correlation coefficients between each dimension's score and the test's total score. The table below illustrates the validity coefficients for the test's dimensions:

Internal Consistency Between Each Dimension's Score and the Total Test Score

Table (3)

Correlation Coefficient Between Each Dimension's Score and the Total Score for the 21st-Century Skills Test (N = 20)

Skills	Critical Thinking	Creativity
Correlation	0.904**	0.983**

**. Correlation is significant at the 0.01 level
The test's dimensions are valid, as shown in the preceding table, which indicates that all correlation coefficients are statistically significant at the 0.01 level.

Discriminant Validity Method

To verify the discriminant validity of the 21st-century skills test, the following procedure was used:

1. High and Low Score Selection: Out of the 20 students in the pilot sample, 27% of the highest and 27% of the lowest scores were selected.
2. Statistical Test: To ascertain the significance of the variations between these two groups' means, the non-parametric Mann-Whitney U Test was employed.

The findings of the z-value and mean rank discrepancies between the two groups are shown in the table below. The following were the results:

Table (4)

Results of Differences Between Mean Ranks and z-Value Between the Two Groups for the 21st-Century Skills Test

Skills	Group	N	Mean Rank	Sum of Ranks	Z- Value	Sig

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

Critical Thinking	The high level	6	9.50	57.00	3.035	0.01
	The low level	6	3.50	21.00		
Creativity	The high level	6	9.50	57.00	2.918	0.01
	The low level	6	3.50	21.00		
All Over The Test	The high level	**6**	**9.50**	**57.00**	**2.918**	**0.01**
	The low level	**6**	**3.50**	**21.00**		

The table shows a statistically significant difference between the two levels at the 0.01 level, indicating a high degree of discriminant validity for the test.

Test Reliability Calculation

The reliability of the test was calculated using the following methods:

1. Cronbach's Alpha Coefficient Method

The test's Cronbach's Alpha coefficient was computed for both the test and its primary and subskills. The following table displays the findings:

Table (5)

Cronbach's Alpha Coefficient for the critical thinking and creativity Skills Test (N = 20)

Skills	Critical Thinking	Creativity	ALL Over The test
Cronbach's Alpha	0.704	0.906	0.885

The previous table shows that the test has a high degree of reliability, as evidenced by the Cronbach's Alpha coefficient.

2. Test-Retest Reliability Method

Students in the pilot sample were given the test, then two weeks later, the same sample was given the test again. Using SPSS (V.18), the Pearson correlation coefficient was calculated for the students' scores across the two administrations. The following table displays the correlation coefficient value between the two administrations for each of the test's competencies and the entire test:

Table (6)

Reliability Coefficient for the 21st-Century Skills Test (N = 20)

Skills	Critical Thinking	Creativity	ALL Over The test
Correlation	0.767**	0.713**	0.894**

The preceding table shows that all correlation coefficients are large and statistically significant at the 0.01 level, demonstrating the test's and each primary skill's reliability.

Calculating Test Duration

The mean time per sample member was used to determine the test duration. Based on this, the researcher determined that an hour is the ideal test duration.

Study Sample Selection and Division into Two Equivalent Groups

Two groups of students from Al Madinah International University were the study's subjects. One was a 17-student experimental group that used [insert experimental method/approach] during study sessions. The other consisted of 17 students in a control group

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

who were instructed using the traditional technique, as indicated in the table below:

Table (6)
Illustrates the Number of Individuals in the Two Study Groups

Group	Control	experimental	Total
Pre	17	17	34
Post	17	17	34

This study is a quasi-experimental study that examines how one or more experimental factors affect one or more dependent factors. Thus, one of the experimental technique designs—more precisely, the pre-test and post-test design for two groups—was used: one for the experimental group and one for the control group.

Experimental Design of the Study

Study Experiment Procedures

Controlling the most important extraneous variables that could affect the dependent variable (21st-century abilities) was crucial to examining the impact of LLMs on it. This control ensures that the independent variable is the only cause of any changes in these variables observed. Among these variables are:

a) **Cultural and Economic Level**: The fact that both research groups were selected from the same social milieu inside Al Madinah International University's administration suggests that their social, cultural, and economic levels are similar. As a result, in this variable, the two groups can be regarded as equal.

b) **Students' Level in 21st-Century Skills**: The Mann-Whitney U Test was employed to ensure the two study groups were equivalent in 21st-century skills. In the pre-application of the 21st-century skills test, this test determined the difference in the mean rankings of students' scores between the experimental and control groups for each of the test's primary and sub-skills. The following table displays this:

Table (7)
Mann-Whitney Test between the Mean Rank Scores of the Experimental and Control Groups of [Insert Variable Name Here] (N1 = N2 = 17)

	Main Skills	Sub Skills	Group	Mean Rank	Sum of Ranks	U-value	Z-value	α Sig
Critical Thinking	Analysing and Evaluating Information	Sub1	Experimental	18.50	314.50	127.50	0.797	**0.426**
			Control	16.50	280.50			
		Sub2	Experimental	17.00	289.00	136.00	0.418	**0.676**
			Control	18.00	306.00			
		Sub3	Experimental	18.50	314.50	127.50	0.886	**0.375**
			Control	16.50	280.50			
	Analysing and Evaluating Information		**Experimental**	**17.79**	**302.50**	**139.50**	**0.192**	0.848
			Control	**17.21**	**292.50**			
	Problem-Solving and Decision-Making	Sub4	Experimental	18.50	315.50	127.50	0.886	**0.375**
			Control	16.50	280.50			
		Sub5	Experimental	18.00	306.00	136.00	0.361	**0.718**

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

			Control	17.00	289.00			
	Problem-Solving and Decision-Making		Experimental	18.71	318.00	124.00	0.807	**0.420**
			Control	16.29	277.00			
	Critical Thinking		**Experimental**	**18.38**	**312.50**	**129.50**	**0.583**	0.560
			Control	**16.62**	**282.50**			
Creativity	Generating New Ideas and Solutions	Sub6	Experimental	16.00	272.00	119.00	1.083	**0.279**
			Control	19.00	323.00			
		Sub7	Experimental	18.50	314.50	127.50	0.742	**0.458**
			Control	16.50	280.50			
	Generating New Ideas and Solutions		**Experimental**	**17.44**	**296.50**	**143.50**	**0.038**	0.970
			Control	**17.56**	**298.50**			
	Adapting and Iterating	Sub8	Experimental	18.50	314.50	127.50	0.742	**0.458**
			Control	16.50	280.50			
		Sub9	Experimental	16.00	272.00	119.00	1.043	**0.297**
			Control	19.00	323.00			
	Adapting and Iterating		**Experimental**	**16.91**	**287.50**	**134.50**	**0.376**	0.707
			Control	**18.09**	**307.50**			
	Creativity		**Experimental**	**16.97**	**288.50**	**135.50**	**0.323**	0.747
			Control	**18.03**	**306.50**			
	All over the test		**Experimental**	**18.59**	**306.50**	**135.50**	**0.316**	0.752
			Control	**16.97**	**288.50**			

The table above shows that the "Z" values for each subskill, significant skill, and the test overall are not statistically significant at the $\alpha \leq 0.05$ level. This shows that, before the study was put into place, the two groups were equivalent in terms of 21st-century abilities overall, as well as in their sub-skills and core skills.

Presentation, Analysis, Discussion, and Interpretation of Results:

1- Presentation and Discussion of the Results Related to the First Hypothesis:

To determine whether the study's first hypothesis—that "There is a statistically significant difference between the mean rank scores of the experimental and control groups in the post-test of 21st-century skills as a whole in favour of the experimental group"—is true, the mean ranks of the experimental and control groups in the post-test of 21st-century skills were compared using the Mann-Whitney Test methodology. Furthermore, the experimental treatment's effect size on 21st-century competencies was computed. This is shown in the following table.

Table (8)
Results of the effect size of the experimental treatment on the 21st-century skills (N1 = N2 = 17)

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

Main Skills	**Group**	**Mean Rank**	**Sum of Ranks**	**U-value**	**Z-value**	**α Sig**	(r_{rb})
All Over the Test	Experimental	26.00	442.00	0.00	5.045	0.000	**1**

The previous table clearly shows that the experimental group's mean rank scores on the post-test of 21st-century abilities differ from the control group's at the level of ($\alpha < 0.01$), which is statistically significant. This suggests that the first research hypothesis has been accepted. Furthermore, the experimental treatment's effect size on 21st-century skills hit 1.00, indicating a substantial impact.

2- Presentation and Discussion of the Results Related to the Second Hypothesis:

Using the Wilcoxon Signed Ranks Test, the significance of the differences between the experimental group's mean rank scores in the pre-test and post-test of 21st-century skills as a whole was assessed to test the validity of the study's second hypothesis, which states that "There is a statistically significant difference between the mean rank scores of the experimental group in the pre-test and post-test of 21st-century skills as a whole, in favor of the post-test." Furthermore, the experimental treatment's effect size on 21st-century competencies was computed. This is shown in the following table.

Table (9)
The differences between the mean rank scores of the experimental group in the pre-test and post-test of 21st-century skills
(N = 17)

Main Skills	Ranks	N	Mean Rank	Sum of Ranks	Z-value	α Sig	(r_{rb})
All Over the Test	Negative Ranks	0	0.00	0.00	3.656	0.01	1
	Positive Ranks	17	9.00	153.00			
	Ties	0					

It is evident from the above table that there is a statistically significant difference at the level of ($\alpha \leq 0.01$) between the mean rank scores of the experimental group in the pre-test and post-test of 21st-century skills, in favour of the post-test. This indicates acceptance of the second research hypothesis. Moreover, the **effect size** of the experimental treatment on 21st-century skills reached 1.00**,** indicating a powerful **impact** of the experimental therapy on these skills.

3- Presentation and Discussion of the Results Related to the Third Hypothesis:

Using the Mann-Whitney Test, the difference between the experimental and control groups' mean rank scores in the post-test of overall critical thinking skills and its main and sub-skills was calculated to test the validity of the study's third hypothesis, which states that: "There is a statistically significant difference between the mean rank scores of the experimental and control groups in the post-test of overall critical thinking skills and its main and sub-skills, in favour of the experimental group." Furthermore, the experimental treatment's effect size on critical thinking abilities was computed. This is shown in the following table.

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

Table (10)
The difference between the mean rank scores of the experimental and control groups in the post-test of overall critical thinking skills and its main and sub-skills
(N1 = N2 = 17)

Main Skills	Sub Skills	Group	**Mean Rank**	**Sum of Ranks**	**U-value**	**Z-value**	**α Sig**	(r_{rb})
Analysing and Evaluating Information	Sub1	Experimental	23.50	399.50	42.50	4.120	0.01	**0.706**
		Control	11.50	195.50				
	Sub2	Experimental	22.50	382.50	59.50	3.433	0.01	**0.588**
		Control	12.50	212.50				
	Sub3	Experimental	24.50	416.50	25.50	4.806	0.01	**0.824**
		Control	10.50	178.50				
Analysing and Evaluating Information		Experimental	25.62	435.50	6.50	5.080	0.01	**0.955**
		Control	9.38	159.50				
Problem-Solving and Decision-Making	Sub4	Experimental	23.00	391.00	51.00	3.743	0.01	**0.647**
		Control	12.00	204.00				
	Sub5	Experimental	22.50	382.50	59.50	3.536	0.01	**0.588**
		Control	12.50	212.50				
Problem-Solving and Decision-Making		Experimental	24.71	420.00	22.00	4.529	0.01	**0.848**
		Control	10.29	175.00				
Critical Thinking		Experimental	26.00	442.00	0.00	5.177	0.01	**1.00**
		Control	9.00	153.00				

The aforementioned table clearly shows that the experimental group's mean rank scores on the post-test of overall critical thinking skills and its major and sub-skills differ from the control group's at the level of ($\alpha \leq 0.01$), indicating a statistically significant difference. This suggests that the third research hypothesis has been accepted. Furthermore, the experimental treatment's effect sizes on critical thinking abilities and primary and secondary skills ranged from moderate to large, suggesting a significant impact.

4- Presentation and Discussion of the Results Related to the Fourth Hypothesis:

Table (11)

The Wilcoxon Signed Ranks Test was used to assess the significance of differences in the experimental group's mean rank scores between the pre-test and post-test for overall critical thinking skills and their main and sub-skills. This was done to test the validity of the study's fourth hypothesis, which states that: "There is a statistically significant difference between the mean rank scores of the experimental group in the pre-test and post-test of overall critical thinking skills and its main and sub-skills, in favour of the post-test." Calculations were made to determine the experimental treatment's effect size on critical thinking abilities. This is shown in the following table.

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

Results of the effect size of the experimental treatment on critical thinking skills (N = 17)

Main Skills	Sub Skills	**Ranks**	**N**	**Mean Rank**	**Sum of Ranks**	**Z-value**	**α Sig**	(r_{rb})
Analysing and Evaluating Information	Sub1	Negative Ranks	0	0.00	0.00	3.317	0.01	**0.137**
		Positive Ranks	11	6.00	66.00			
		Ties	6					
	Sub2	Negative Ranks	0	0.00	0.00	3.464	0.01	**0.146**
		Positive Ranks	12	6.50	78.00			
		Ties	5					
	Sub3	Negative Ranks	0	0.00	0.00	3.606	0.01	**0.189**
		Positive Ranks	13	7.00	91.00			
		Ties	4					
Analysing and Evaluating Information		Negative Ranks	0	0.00	0.00	3.573	0.01	**0.778**
		Positive Ranks	16	8.50	136.00			
		Ties	1					
Problem-Solving and Decision-Making	Sub4	Negative Ranks	0	0.00	0.00	3.317	0.01	**0.137**
		Positive Ranks	11	6.00	66.00			
		Ties	6					
	Sub5	**Negative Ranks**	**0**	**0.00**	**0.00**	**3.162**	**0.01**	0.112
		Positive Ranks	**10**	**5.50**	**55.00**			
		Ties	**7**					
Problem-Solving and Decision-Making		**Negative Ranks**	**0**	**0.00**	**0.00**	3.391	0.01	**0.373**
		Positive Ranks	**14**	**7.50**	**105.00**			
		Ties	**3**					
Critical Thinking		**Negative Ranks**	**0**	**0.00**	**0.00**	3.662	0.01	**1**
		Positive Ranks	**17**	**9.00**	**153.00**			
		Ties	**0**					

The preceding table makes it clear that the experimental group's mean rank scores on the pre-test and post-test of overall critical thinking skills and its main and sub-skills differ significantly at the level of ($\alpha < 0.01$), favouring the post-test. This suggests that the fourth research hypothesis has been accepted. The experimental treatment had a substantial effect size on the ability to analyse and evaluate information as well as on general critical thinking skills, but a poor effect size on the other skills.

5- Presentation and Discussion of the Results Related to the Fifth Hypothesis:

The Mann-Whitney Test was used to determine differences in mean rank scores between the experimental and control groups in the post-test for overall creativity and its main and sub-skills.

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

This was done to test the validity of the study's fifth hypothesis, which states that: "There is a statistically significant difference between the mean rank scores of the experimental and control groups in the post-test of overall creativity and its main and sub-skills, in favour of the experimental group." Furthermore, the experimental treatment's effect size on creativity skills was computed. This is shown in the following table.

Table (12)
The effect size of the experimental treatment on creativity skills (N1 = N2 = 17)

Main Skills	Sub Skills	Group	**Mean Rank**	**Sum of Ranks**	**U-value**	**Z-value**	**α Sig**	(r_{rb})
Generating New Ideas and Solutions	Sub6	Experimental	24.35	414.00	28.00	4.255	0.01	**0.806**
		Control	10.65	181.00				
	Sub7	Experimental	24.76	421.00	21.00	4.480	0.01	**0.854**
		Control	10.24	174.00				
Generating New Ideas and Solutions		Experimental	25.82	439.00	3.00	5.00	0.01	**0.979**
		Control	9.18	156.00				
Adapting and Iterating	Sub8	Experimental	24.94	424.00	18.00	4.569	0.01	**0.875**
		Control	10.06	171.00				
	Sub9	Experimental	25.12	427.00	15.00	4.645	0.01	**0.896**
		Control	9.88	168.00				
Adapting and Iterating		Experimental	25.94	441.00	1.00	5.068	0.01	**0.993**
		Control	9.06	154.00				
Creativity		Experimental	26.00	442.00	0.00	5.070	0.01	**1.00**
		Control	9.00	153.00				

It is evident from the above table that there is a statistically significant difference at the level of ($\alpha \leq 0.01$) between the mean rank scores of the experimental and control groups in the post-test of overall creativity skills and its main and sub-skills, in favour of the experimental group. This indicates acceptance of the fifth research hypothesis. The effect size of the experimental treatment on creativity skills and their main and sub-skills was large, indicating a substantial impact.

6- Presentation and Discussion of the Results Related to the Sixth Hypothesis:

The Wilcoxon Signed Ranks Test was used to assess the significance of differences in the experimental group's mean rank scores between the pre-test and post-test for overall creativity skills and their main and sub-skills. This was done in order to test the validity of the sixth hypothesis of the study, which states that: "There is a statistically significant difference between the mean rank scores of the experimental group in the pre-test and post-test of overall creativity skills and its main and sub-skills, in favour of the post-

Received Date: 10/06/2026 |Revised Date: 10/07/2026 | Acceptance Date: 10/08/2026 | Published Date: 01/09/2026

test." Furthermore, the experimental treatment's effect size on creativity skills was computed. This is shown in the following table.

Table (13)
The **effect size** *of the experimental treatment on creativity skills (N = 17)*

Main Skills	Sub Skills	**Ranks**	**N**	**Mean Rank**	**Sum of Ranks**	**Z-value**	**α Sig**	(r_{rb})
Generating New Ideas and Solutions	Sub6	Negative Ranks	0	0.00	0.00	3.695	0.01	**1**
		Positive Ranks	17	9.00	153.00			
		Ties	0					
	Sub7	Negative Ranks	0	0.00	0.00	3.695	0.01	**1**
		Positive Ranks	17	9.00	153.00			
		Ties	0					
Generating New Ideas and Solutions		Negative Ranks	0	0.00	0.00	3.671	0.01	**1**
		Positive Ranks	17	9.00	153.00			
		Ties	0					
Adapting and Iterating	Sub8	Negative Ranks	0	0.00	0.00	3.787	0.01	**1**
		Positive Ranks	17	9.00	153.00			
		Ties	0					
	Sub9	**Negative Ranks**	**0**	**0.00**	**0.00**	**3.695**	**0.01**	1
		Positive Ranks	**17**	**9.00**	**153.00**			
		Ties	**0**					
Adapting and Iterating		**Negative Ranks**	**0**	**0.00**	**0.00**	3.684	0.01	**1**
		Positive Ranks	**17**	**9.00**	**153.00**			
		Ties	**0**					
Creativity		**Negative Ranks**	**0**	**0.00**	**0.00**	3.655	0.01	**1**
		Positive Ranks	**17**	**9.00**	**153.00**			
		Ties	**0**					

ISSN 2600-7398 الترقيم الدولي

Received Date: 16/05/2026 |Revised Date: 05/07/2026 | Acceptance Date: 13/08/2026 | Published Date: 01/09/2026

It is evident from the above table that there is a statistically significant difference at the level of ($\alpha \leq 0.01$) between the mean rank scores of the experimental group in the pre-test and post-test of overall creativity skills and its main and sub-skills, in favour of the post-test. This indicates acceptance of the sixth research hypothesis. The effect size of the experimental treatment on creativity skills and all their main and sub-skills was very large, indicating a substantial positive impact of the experimental intervention.

Conclusion

This study shows that Large Language Models (LLMs) can significantly boost critical thinking and creativity among English as a Foreign Language (EFL) learners. The experimental group made impressive gains, and the effect sizes are substantial, highlighting the potential of including LLMs in language teaching. These results help fill a gap in the research by showing that LLMs can improve higher-order skills, which are vital for learners in the twenty-first century, beyond just basic language tasks. The strength of these findings comes from the study's solid methodology, which includes thorough validation and reliability testing of the assessment tool.

This study has far-reaching implications. The findings suggest that LLMs can play a crucial role in a teaching approach that encourages deeper cognitive engagement, rather than simply serving as extra resources for instructors. Incorporating LLMs into the EFL curriculum enables personalised learning experiences, immediate feedback, and interactive practice opportunities that traditional methods may lack. These results support the development of systems to guide the effective use of LLM-based activities and are essential for making strategic investments in these efforts. However, issues such as potential biases in LLM outputs and the need for human oversight still need to be addressed to ensure fair and ethical use.

Future studies should investigate the precise ways in which LLMs affect creativity and critical thinking, perhaps by examining several LLM intervention types and their differing outcomes. To evaluate the long-term effects of LLM integration on skill development, longer-term research might also be beneficial. Furthermore, a more thorough understanding of LLMs' contribution to 21st-century skills would come from examining how they enhance the other two "Cs"—communication and collaboration. In the end, our study contributes to the growing body of research on the careful incorporation of cutting-edge AI technology in the classroom, opening the door to more dynamic, captivating, and productive language-learning environments.

References

Abdali, S., Anarfi, R., Barberan, C. J., & He, J. (2024). Securing large language models: Threats, vulnerabilities and responsible practices. *arXiv preprint arXiv:2403.12503*.

Alfares, N. (2024). Does applying inductive methods based on systems thinking improve EFL grammar learning?. *Saudi Journal of Language Studies*, *4*(3), 162–174.

Alias, S. (2024). Opportunities and Challenges in the Cultivation of Software Development Professionals in the Context of Large Language Models. *Proceedings of the 2024 International Symposium on Artificial Intelligence for Education*. https://doi.org/10.1145/3700297.3700342.

Almukhanbet, S., Moldabek, K., Orynbassarova, G., Yesnazar, A., & Zhorabekova, A. (2025, May). Developing primary school students' language

Received Date: 16/05/2026 |Revised Date: 05/07/2026 | Acceptance Date: 13/08/2026 | Published Date: 01/09/2026

skills from a critical thinking perspective: a methodological system. In *Frontiers in Education* (Vol. 10, p. 1565783). Frontiers Media SA.

Arthars, N., Thompson, K., Huijser, H., Kickbusch, S., Cunningham, S., Winter, G., ... & Lockyer, L. (2024). Formative assessment of group work skills: An analytics-enabled conceptual framework. *Australasian Journal of Educational Technology*, *40*(4), 139–154.

Azzalini, A. (2023). Critical Thinking, Questioning and Reasoning as Inclusive Teaching Methods for English Language Learning with Gifted Students. EL.LE, 3. https://doi.org/10.30687/elle/2280-6792/2023/03/003

Ballard, D. H., Antigua-Made, A., Barre, E., Edney, E., Gordon, E. B., Kelahan, L., ... & Adams, S. J. (2025). Impact of ChatGPT and large language models on radiology education: association of academic radiology—radiology research alliance task force white paper. *Academic radiology*, *32*(5), 3039–3049.

Barzang, E. (2019). Fostering EFL learners' grammar achievement using recasts and meta-linguistic awareness error correction feedback. *Journal for the Study of English Linguistics*, *8*(1), 35.

Behzadpoor, S., Fathi, J., & Greenier, V. (2023). Teaching for creativity in an EFL context: The predictive roles of school climate, teaching enthusiasm, and metacognition. *Thinking Skills and Creativity*. https://doi.org/10.1016/j.tsc.2023.101419.

Chiang, F. K., Zhu, D., & Yu, W. (2022). A systematic review of academic dishonesty in online learning environments. *Journal of Computer Assisted Learning*, *38*(4), 907–928.

Dan, Q., Bai, B., & Huang, Q. (2024). Gender differences in the relations between EFL students' classroom relationships and English language proficiency: The mediating role of self-regulated learning strategy use. *System*, *123*, 103311.

Darwin, R. D., M., N., Suryati, N. L., E. D., & Marzuki. (2024). Critical thinking in the AI era: An exploration of EFL students' perceptions, benefits, and limitations. *Cogent Education*, *11*(1), 2290342.

Ding, Q., Cao, H., Cao, Z., & Zhao, T. (2025). Dual-perspective fusion for word translation enhancement. *Information Fusion*, *117*, 102815.

Do, H. N., Do, B. N., & Nguyen, M. H. (2023). 3How do constructivism learning environments generate better motivation and learning strategies? The Design Science Approach. *Heliyon*, *9*(12).

Eto, H. (2022). Aspects in the Development of Critical Thinking in Asian EFL Higher Education: A Critical Review. *International Journal of Linguistics, Literature and Culture*, *9*(4), 1-15.

Ferdoush, Z., & Jahan, R. (2024). Exploring the Challenges Encountered by EFL Teachers in Promoting Critical Thinking Skills. International Journal of Research and Innovation in Social Science, VIII(IIIS), 2053–2062. https://doi.org/10.47772/ijriss.2024.803148s

Goldie, J. G. S. (2016). Connectivism: A knowledge learning theory for the digital age?. *Medical teacher*, *38*(10), 1064-1069.

Gouloudis, S. E. (2020). Interaction as a Method of Educating Constructivism: The Gaming-Based Learning Concept as a Tool of Interactivity. In *Advanced Technologies and Standards for Interactive Educational Television: Emerging Research and Opportunities* (pp. 170-205). IGI Global Scientific Publishing.

Guzmán-Simón, F., Torres-Gordillo, J. J., & Caballero, K. (2020). Understanding assessment processes for communicative competence through an

Received Date: 16/05/2026 |Revised Date: 05/07/2026 | Acceptance Date: 13/08/2026 | Published Date: 01/09/2026

analysis of teachers' reported practice. *Education Sciences*, *10*(4), 116.

Haltaufderheide, J., & Ranisch, R. (2024). *The ethics of ChatGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs). npj Digital Medicine, 7 (1), 183*.

Hassan Khalil, H. (2015). Using online communication tools in group activities to develop collaborative learning skills. Conference on Education 2015. Partners in Excellence. Conference on Education 2015. Partners in Excellence. https://doi.org/10.5339/qproc.2015.coe.23

Hendricks, G. P. (2019). Connectivism as a Learning Theory and Its Relation to Open Distance Education. Progressio, 41(1). https://doi.org/10.25159/2663-5895/4773

Hwang, G. J., Rahimi, M., & Fathi, J. (2024). Enhancing EFL learners' speaking skills, foreign language enjoyment, and language-specific grit utilising the affordances of a MALL app: A microgenetic perspective. *Computers & Education*, *214*, 105015.

Jabeen, M. R., & Sarifa, N. (2024). Conceptualizing Creativity: Leading to an Efficacious EFL Teaching. *European Journal of English Language and Literature Studies*, *12*(7), 41-57.

Jeny, D. P. (2024). Classrooms to Networks: Applying Connectivism Principles in Pedagogy``. *Journal of Pedagogi*, *1*(5), 63-73.

Karusala, N., Holeman, I., & Anderson, R. (2019). Engaging identity, assets, and constraints in designing for resilience. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1-23.

Kim, M.-S. (2022). A Case Study of a Connectivism-Based Lecture as an Online Teaching and Learning Course in the Digital Era. The Korean Association of General Education, 16(1), 131–146. https://doi.org/10.46392/kjge.2022.16.1.131

Koga, S. (2023). Exploring the pitfalls of large language models: Inconsistency and inaccuracy in answering pathology board examination-style questions. *Pathology International*, 73. https://doi.org/10.1111/pin.13382.

Kumar, P. (2024). Large language models (LLMs): survey, technical frameworks, and future challenges. *Artificial Intelligence Review*, *57*(10), 260.

Lan, J. (2024). The Promotion of Students' Critical Thinking Skills through the Integration of English Civics Elements into EFL Classrooms. Applied Mathematics and Nonlinear Sciences, 9(1). https://doi.org/10.2478/amns-2024-2203

Li, C., & Wei, L. (2025). The enhancement of creativity through foreign language learning: Do personality traits matter?. *Bilingualism: Language and Cognition*, *28*(2), 510-521.

Li, R., Pandurangan, P., Frluckaj, H., & Dabbish, L. (2021). Code of conduct conversations in open source software projects on github. *Proceedings of the ACM on Human-computer Interaction*, *5*(CSCW1), 1-31.

Lin, Z., Guan, S., Zhang, W., Zhang, H., Li, Y., & Zhang, H. (2024). Towards trustworthy LLMs: a review on debiasing and dehallucinating in large language models. *Artificial Intelligence Review*, *57*(9), 243.

Linders, J., & Tomczak, J. M. (2025). Knowledge graph-extended retrieval augmented generation for question answering. *arXiv preprint arXiv:2504.08893*.

Madi, M., Alhumaid, J., Asiri, A., & Sadaf, S. (2024). Empowering dental students' collaborative learning using peer assessment. *Medical Education*, 59, 238 - 239. https://doi.org/10.1111/medu.15580.

Mahowald, K., Tenenbaum, J., Blank, I., Fedorenko, E., Ivanova, A., & Kanwisher, N. (2024).

Received Date: 16/05/2026 |Revised Date: 05/07/2026 | Acceptance Date: 13/08/2026 | Published Date: 01/09/2026

Dissociating language and thought in large language models. *Trends in Cognitive Sciences*, 28, 517-540.https://doi.org/10.1016/j.tics.2024.01.011.

Mehta, S. R., & Al Mahrooqi, R. (2024). Ensuring Inclusivity Through Critical Thinking In EFL Contexts. *Journal of Teaching English for Specific and Academic Purposes*, 025-038.

Meyer, A., Kleinknecht, M., & Richter, D. (2023). What makes online professional development effective? The effect of quality characteristics on teachers' satisfaction and changes in their professional practices. Computers &Education, 200, 104805. https://doi.org/10.1016/j.compedu.2023.104805

Mukta, M., Mim, M., Sakib, S., Fahad, N., Raiaan, M., Ali, M., Ahmad, J., Azam, S., & Fatema, K. (2024). A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges. *IEEE Access*, 12, 26839-26874. https://doi.org/10.1109/ACCESS.2024.3365742.

Mzeil, A. H. (2024). The Significance of Integrating Critical Thinking Perspectives in EFL Programs. *Journal of Ecohumanism*, *3*(4), 3265-3273.

Negoescu, A. G. (2023). The value of critical thinking in the language classroom. *Land Forces Acad. Rev*, *28*, 303-308.

Nguyen, P. H. (2024). Challenges Vietnamese EFL students faced when learning and meeting the requirements of B1 English language proficiency. *Journal of Language Teaching and Research*, *15*(3), 893-901.

Nithideechaiwarachok, B., Chano, J., & Cheng Wu, C. (2024). Socio-cultural and Social Constructivist Theories and Its Application in EFL Classroom for Thai Pre-service Teachers: A Review for Further Research. *International Journal of Language Education*, *8*(3). https://doi.org/10.26858/ijole.v8i3.66499

Nurhidayat, E., Mujiyanto, J., Yuliasri, I., & Hartono, R. (2024). Technology integration and teachers' competency in the development of 21st-century learning in EFL classroom. *Journal of Education and Learning (EduLearn)*, *18*(2), 342-349.

Oroujlou, N., & Sadeghi, K. (1999). Effects of explicit teaching of critical thinking strategies on EFL learners' reading comprehension. *Teaching English Language*, *16*(2), 1-29.

Patil, R., & Gudivada, V. (2024). A Review of Current Trends, Techniques, and Challenges in Large Language Models (LLMs). *Applied Sciences*. https://doi.org/10.3390/app14052074.

Peng, F. (2023). RETRACTED ARTICLE: Evaluating critical thinking of English learners using modern technologies and GTMA. Soft Computing, 28(S2), 513–513. https://doi.org/10.1007/s00500-023-08127-7

Presdyasmara, C. A. R., & Anam, S. U. (2023). Investigating Critical Thinking Skills Of Senior High School Students Through Debate In EFL Classroom. *Journal Education And Development*, *11*(3), 192-198.

Rawlings, B. S., Chetwynd-Talbot, D., Husband, E., Nuttall, A., Quinn, E., Taggart, R., & Roome, H. E. (2025). Divergent thinking is linked with convergent thinking, implications for models of creativity. *Thinking & Reasoning*, 1-23.

Rungsawang, S., & Pookcharoen, S. (2023). Exploring EFL University Teachers' Technology Integration to Promote Thai EFL University Students' Critical Thinking Skills. *English Language Teaching*, *16*(9).

San Jose, D. B. S., Guitche, M. Q., & Madrigal, D. V. (2024). English Proficiency of Japanese Third-Age Learners in the English Foreign Language Program: An Explanatory-Sequential

Received Date: 16/05/2026 |Revised Date: 05/07/2026 | Acceptance Date: 13/08/2026 | Published Date: 01/09/2026

Inquiry. *3L, Language, Linguistics, Literature*, *30*(4), 220-240.

Sinha, N. (2025). QA-prompting: Improving Summarization with Large Language Models using Question-Answering. *arXiv preprint arXiv:2505.14347*.

Song, X., Razali, A. B., Sulaiman, T., Jeyaraj, J. J., & Ds, P. (2024). Impact of project-based learning on critical thinking skills and language skills in EFL context: A review of literature. *World*, *14*(5).

Su, H. (2025). Critical thinking-embedded EFL translation instruction and its impact on EFL translation learners. *Thinking Skills and Creativity*, *57*, 101827.

Tawafak, R. M., Al-Obaydi, L. H., Klimova, B., & Pikhart, M. (2023). Technology integration of using digital gameplay for enhancing EFL college students' behavior intention. *Contemporary Educational Technology, 15*(4), ep452. https://doi.org/10.30935/cedtech/13454

Thanasi-Boçe, M., & Hoxha, J. (2024). From ideas to ventures: building entrepreneurship knowledge with LLM, prompt engineering, and conversational agents. *Education and Information Technologies*, 1-57.

Tin, T. B. (2024). Finding constraints to foster creativity in language learning tasks: An autoethnographic approach. *TESOL Journal*, *15*(3), e792.

Todorovska, R. (2024). Critical thinking in the language classroom – an important 21st century skill. International Journal of Education and Philology, 5(1), 39–48. https://doi.org/10.69648/bfsi1617

Vendri, R., & Jamaris, J. (2024). Constructivism Thinking and Implementation in Non-Formal Education. *SPEKTRUM: Jurnal Pendidikan Luar Sekolah (PLS)*, *12*(3), 450-456.

Wang, H. (2024). Exploring high school teachers' and students' perceptions of teaching for creativity in EFL classes. *Innovation in Language Learning and Teaching*. https://doi.org/10.1080/17501229.2024.2388631.

Wangmanee, P., & Kaowiwattanakul, S. (2024). The Effect of Using the Contemplative Approach to Enhance Speaking and Critical Thinking Skills in the EFL Undergraduate Literature Classroom. *Theory and Practice in Language Studies*, *14*(4), 966-975.

Yeh, H., & Chen, C. (2024). Scripted synergy: elevating EFL writing and creativity through collaborative digital storytelling. *Technology, Pedagogy and Education*, 34, 91 - 104. https://doi.org/10.1080/1475939X.2024.2382970.

Zeng, G., Xue, Q., & Lu, X. (2024, August). Instruction Tuning with LLMs for Programming Exercise Generation. In *International Conference on Web Information Systems and Applications* (pp. 377-385). Singapore: Springer Nature Singapore.

Zhou, Y. (2025). Slide2Text: Leveraging LLMs for Personalized Textbook Generation from PowerPoint Presentations. *arXiv preprint arXiv:2503.17710*.