Journal of Linguistic and Language Education in Technology
https://ejournal.lp3msakti.com/index.php/JLLET/
Email: journaljllet@gmail.com
DOI: https://doi.org/10.36312/jllet.v1i2.72

November 2026, Vol. 1, No. 2
e-ISSN: 3164-7499
pp. 82-101

From Prompt to Argument: A Qualitative Study of EFL Learners' Critical Reasoning in Gemini-Assisted Genre-Based Writing

[1]* Hussien Jawad Abdulridha, [2]Anwar Zulkarnaen, [3]Eliano Zaretsky

[1] Department of English, College of Education, Al-Zahraa University for Women, Karbala-Baghdad Road (opposite pole 70) in the Holy Governorate of Karbala, Iraq

[2] School of Education, Universiti Utara Malaysia, Sintok, Kedah Darul Aman, Malaysia

[3]Faculty of Education, University of Pretoria, Pretoria, South Africa

*Corresponding Author Email: hussien.jawad@alzahraa.edu.iq

Received: September 2026, Revised: October 2026; Accepted: October 2026; Published: November 2026

Abstract

The increasing use of generative artificial intelligence in English as a Foreign Language (EFL) writing raises questions about whether learners critically evaluate AI-generated information or accept linguistically fluent outputs without sufficient scrutiny. This qualitative descriptive case study investigated how EFL learners critically evaluated information and arguments generated by Gemini during genre-based academic writing and how genre-based learning supported such reasoning. Seventeen EFL learners from private-school educational settings in Iraq were purposively selected based on their experience with online learning and academic writing. Data were generated through semi-structured interviews and classroom observations and analyzed iteratively through data condensation, data display, and conclusion drawing and verification. The findings revealed four interconnected dimensions of critical evaluation: assessing the credibility and accuracy of Gemini-generated information, evaluating evidence and argument strength, identifying assumptions and overgeneralizations, and judging genre appropriateness when deciding whether content should be accepted, rejected, or revised. Genre-based learning further supported these practices through building knowledge of the field, analyzing model texts, joint construction, independent text production, and reflection and revision. These stages provided conceptual, rhetorical, collaborative, and metacognitive scaffolding that shifted learners from passive acceptance toward selective and reasoned integration of AI-generated material. The study concludes that Gemini can contribute to academic writing when its outputs are treated as objects of critical inquiry rather than authoritative answers. Integrating explicit verification, evidence evaluation, rhetorical analysis, and reflective decision-making into genre-based pedagogy can help preserve learner agency and promote responsible human–AI collaboration in EFL academic writing.

Keywords: EFL writing; Critical reasoning; Gemini; Generative artificial intelligence; Genre-based pedagogy

Journal of Linguistic Metacognition and Language Learning is Licensed Under a CC BY-SA Creative Commons Attribution-ShareAlike 4.0 International License.

How to cite: Abdulridha, H.J., Zulkarnaen, A., & Zaretsky, E. (2026). From Prompt to Argument: A Qualitative Study of EFL Learners' Critical Reasoning in Gemini-Assisted Genre-Based Writing, *Journal of Linguistic and Language Education in Technology*, 1(2), 82-101, Doi: https://doi.org/10.36312/jllet.v1i2.72

INTRODUCTION

The rapid development of generative artificial intelligence (GenAI) is reshaping academic writing practices in higher education, particularly in English as a Foreign Language (EFL) contexts. Tools such as ChatGPT, Gemini, Claude, and DeepL Write can generate ideas, model academic texts, suggest arguments, reorganize information, and provide immediate feedback on grammar, vocabulary, cohesion, and sentence structure. These affordances are especially attractive to EFL learners, who often need to manage linguistic accuracy and academic conventions simultaneously during writing. Recent research consequently reports considerable potential for GenAI to facilitate language development, drafting, feedback, and

writing efficiency (Li, 2024; Ma & Vasinda, 2025; Mi et al., 2025). Studies also indicate that appropriately integrated AI applications may contribute to learners' cognitive and metacognitive strategies and writing achievement (Chen & Alibakhshi, 2025). At the same time, the rapid expansion of GenAI research in EFL education has generated increasing concern about how learners engage cognitively with AI-generated information rather than simply whether they use AI effectively (Chen, 2025; Xiao & Zhu, 2025). Accordingly, the educational challenge has shifted from providing access to GenAI toward developing learners who can use these technologies critically, responsibly, and purposefully.

This concern is particularly important because linguistic sophistication in AI-generated text does not necessarily indicate strong reasoning or reliable knowledge. GenAI can produce fluent, cohesive, and academically styled responses while simultaneously presenting unsupported claims, questionable evidence, inappropriate generalizations, or weaknesses in argumentation. Research on AI-supported writing therefore highlights an important distinction between improvements in surface-level language quality and the development of higher-order intellectual competencies (Faisal, 2026; Hao et al., 2026; Ma & Vasinda, 2025; Tang, 2024). For EFL learners, this distinction may be especially consequential because their attention during writing can be directed toward grammatical correctness, lexical sophistication, and textual fluency. A response that appears linguistically authoritative may consequently be accepted without sufficient interrogation of its factual accuracy, evidential adequacy, or logical coherence. When students copy, paraphrase, or minimally modify AI-generated content, writing may become more polished while opportunities to evaluate claims, examine assumptions, synthesize evidence, and construct independent arguments are reduced. Thus, the central concern is not simply AI dependence but the possibility of cognitive offloading, in which important reasoning processes are transferred from learners to the technology (Faisal, 2026; Tang, 2024; Yang et al., 2025).

These developments make critical reasoning a central component of GenAI-assisted EFL writing. Critical reasoning in this context extends beyond identifying obviously incorrect information. It involves evaluating the credibility and relevance of claims, distinguishing assertions from evidence, identifying assumptions and contradictions, assessing evidentiary strength, comparing alternative perspectives, and determining whether conclusions follow logically from supporting information. Such abilities increasingly form part of GenAI literacy, which requires learners not merely to operate AI tools but to understand their affordances and limitations, evaluate their outputs, and make informed ethical decisions about their use (Jiang, 2026). Academic integrity further complicates this process because AI-assisted writing raises questions regarding originality, attribution, authorship, transparency, and responsibility for generated content (Han, 2025; Mohammadkarimi, 2023). Consequently, critical reasoning should not be treated as an incidental outcome of AI-supported writing. It needs to become an explicit instructional goal that enables learners to decide what AI-generated information should be accepted, rejected, verified, modified, or supported with additional evidence.

Genre-based pedagogy (GBP) provides a potentially productive framework for achieving this goal. Rather than treating writing primarily as the production of grammatically correct sentences, GBP foregrounds the social purpose, audience, rhetorical organization, and disciplinary conventions of particular genres. These characteristics provide concrete criteria against which AI-generated responses can be examined. Within genre-based learning, learners can evaluate whether a Gemini-generated claim serves the communicative purpose of a text, whether supporting evidence is appropriate, whether rhetorical moves are logically organized, and whether the response conforms to disciplinary expectations (Chen, 2025; Jiang, 2026; Zhang & Samsudin, 2026). The pedagogical stages commonly associated with genre-based learning—building knowledge of the field, analyzing model texts, jointly constructing texts, and independently producing texts—also create opportunities for progressively developing

critical engagement with AI. Rather than positioning Gemini as an authoritative source or automatic text producer, the technology can therefore become an object of inquiry whose output must be examined against explicit genre criteria.

Such an approach changes the pedagogical role of GenAI in academic writing. During knowledge building, students can compare AI-generated information with scholarly sources and identify unsupported or inaccurate claims. During text modeling, they can examine whether Gemini reproduces the rhetorical moves expected in a particular academic genre. During collaborative construction, students can discuss alternative interpretations, evaluate the adequacy of evidence, and justify modifications to AI-generated arguments. During independent writing, learners can be required to demonstrate why specific AI suggestions were retained, rejected, or revised. Explicit evaluative checkpoints and rubrics may further direct attention toward argument quality, evidence, coherence, source credibility, and genre appropriateness rather than linguistic correctness alone (Divya & Alamelu, 2026; Roe et al., 2026). Metacognitive reflection can additionally encourage learners to monitor how they interact with AI, question the reliability of its outputs, and evaluate whether reliance on AI enhances or substitutes for their own reasoning (Chen & Alibakhshi, 2025; Tang, 2024). In this way, genre-based pedagogy can connect AI literacy with observable reasoning practices during the writing process.

Nevertheless, existing scholarship has not yet sufficiently clarified how this integration develops learners' critical reasoning. Much research on GenAI-assisted L2 and EFL writing has concentrated on writing performance, linguistic improvement, learner perceptions, feedback, technology acceptance, or broad AI literacy (Li, 2024; Mi et al., 2025; Wang et al., 2025; Xiao & Zhu, 2025). Studies examining higher-order thinking suggest positive possibilities but also show that such outcomes are unlikely to emerge automatically from access to AI; they depend substantially on instructional design and explicit cognitive scaffolding (Hao et al., 2026; Tang, 2024). Recent systematic analyses similarly identify the need for theoretically grounded investigations into how GenAI can support rather than displace cognitive engagement in L2 writing (Tang et al., 2026; Yang et al., 2025). Although AI-integrated genre pedagogy has begun to receive attention, current work has focused notably on teachers' AI literacy and pedagogical competence (Zhang & Samsudin, 2026). Consequently, limited attention has been directed specifically to how EFL learners critically evaluate the information and arguments generated by a particular GenAI system such as Gemini while completing genre-based academic writing tasks.

This gap is important because different forms of AI-supported writing may create different cognitive demands. Using AI for grammar correction is fundamentally different from using it to generate claims, evidence, explanations, or complete argumentative passages. The latter directly affects the epistemic dimension of academic writing because students must judge what constitutes credible knowledge and defensible argumentation. Furthermore, learner proficiency, genre, disciplinary context, teacher mediation, and task design can influence how critically students engage with AI output (Faisal, 2026; Mi et al., 2025; Yang et al., 2025). Therefore, evidence that GenAI improves writing quality cannot by itself demonstrate that learners are becoming more critical writers. Research needs to investigate the reasoning processes through which learners interrogate AI-generated content and determine how genre-based scaffolding can promote those processes.

Against this background, the novelty of the present study lies in positioning Gemini-generated information and arguments not merely as writing assistance but as material for systematic critical interrogation within genre-based learning. Rather than measuring GenAI effectiveness primarily through linguistic improvement or perceptions of usefulness, the study focuses on learners' critical reasoning when evaluating AI-generated claims, evidence, assumptions, logical relationships, and genre appropriateness. It therefore connects three

domains that have often been examined separately: Gemini-assisted writing, EFL learners' critical reasoning, and genre-based pedagogy. This integration is expected to provide a more cognitively oriented account of human–AI collaboration in academic writing and to clarify how structured genre pedagogy can prevent fluent AI-generated language from substituting for students' independent judgment.

Accordingly, this study aims (1) to investigate how EFL learners critically evaluate information and arguments generated by Gemini during genre-based academic writing tasks, and (2) to examine how genre-based learning facilitates the development and enactment of critical reasoning in learners' interactions with Gemini-generated content. These objectives are addressed through the following research questions: RQ1: How do EFL learners critically evaluate the information and arguments generated by Gemini during genre-based academic writing tasks? RQ2: How does genre-based learning support EFL learners' critical reasoning when they evaluate, select, revise, and integrate Gemini-generated content into their academic writing?

RESEARCH METHOD

Research Design

This study employed a qualitative descriptive case-study design to investigate EFL learners' critical reasoning when engaging with Gemini-generated information and arguments during genre-based academic writing activities. A qualitative approach was considered appropriate because the study sought to understand learners' reasoning processes, judgments, interpretations, and decision-making practices as they interacted with generative artificial intelligence rather than to measure predetermined variables or establish causal relationships. More specifically, the study focused on how learners evaluated AI-generated claims, assessed the adequacy of supporting evidence, identified questionable or unsupported information, and decided whether particular Gemini-generated content should be accepted, rejected, verified, or revised for inclusion in their academic writing.

The case-study orientation enabled the phenomenon to be examined within its authentic instructional context: an online EFL academic-writing classroom incorporating Gemini and genre-based learning. This design was particularly suitable for addressing the first research question, which examined how EFL learners critically evaluated information and arguments generated by Gemini, because such evaluation involves processes that cannot be adequately understood through writing scores alone. Interviews allowed learners to explain the reasoning underlying their decisions, whereas classroom observation provided evidence of how those decisions were enacted during actual writing activities. The design was also appropriate for the second research question concerning how genre-based learning supported learners' critical reasoning. Observation of genre-based activities made it possible to examine critical reasoning across instructional stages, including engagement with genre models, evaluation of arguments, collaborative discussion, revision, and independent text construction.

The study therefore conceptualized critical reasoning as an observable and explainable process rather than simply as a final writing outcome. Particular attention was given to behaviors such as questioning the credibility of Gemini-generated claims, distinguishing evidence from unsupported assertions, checking information against other sources, identifying inconsistencies, considering alternative interpretations, examining logical relationships between claims and evidence, and evaluating whether AI-generated material conformed to the communicative purpose and rhetorical conventions of the target genre. This contextualized approach was consistent with the broader objective of investigating the relationship between AI-supported writing and learners' active intellectual engagement.

One limitation of the qualitative case-study design is that findings from a small and context-specific group cannot be statistically generalized to all EFL learners. The findings

instead provide analytical and contextual insights into how critical reasoning operates within a particular educational environment. Another potential limitation concerns researcher interpretation. Decisions about whether a learner's behavior represented verification, evaluation, acceptance, rejection, or revision necessarily involved interpretive judgment. To minimize this risk, the analysis was guided by clearly defined analytical categories derived from the research questions and the conceptualization of critical reasoning. Evidence from interviews was also compared with observational data so that interpretations were not based on a single source.

Participant reactivity represented another potential source of bias because students might modify their behavior when they knew that their classroom interaction with Gemini was being observed. This was mitigated by conducting observations during regular learning activities rather than introducing artificial experimental tasks wherever possible. In addition, observations focused on patterns of interaction and reasoning rather than evaluating individual students' performance. Social-desirability bias was also considered during interviews, particularly because students might present themselves as more critical or responsible users of AI than they were in practice. Triangulating interview accounts with classroom observations helped identify consistencies and discrepancies between what participants reported and what they actually did during Gemini-assisted writing activities. These procedures strengthened the credibility and dependability of the qualitative interpretation.

Participants

The participants were 17 EFL learners studying in private-school educational settings in Iraq. Purposive sampling was employed because the study required participants with experience relevant to the phenomenon being investigated. Rather than selecting learners randomly, participants were chosen according to their familiarity with online learning, academic writing activities, and the use of digital learning environments. These criteria were important because the study involved interaction with Gemini within an online genre-based academic-writing context and therefore required participants who were sufficiently comfortable working in digitally mediated instructional environments.

The principal inclusion criteria were that participants: (1) were EFL learners enrolled in a private-school context in Iraq; (2) had previous experience participating in online learning activities; (3) had experience completing academic or structured English writing tasks; (4) were able to participate throughout the Gemini-assisted genre-based writing activities; and (5) agreed voluntarily to participate in interviews and classroom observations. Previous expertise with Gemini was not treated as evidence of critical reasoning competence. Familiarity with GenAI was regarded only as contextual information because frequent use of AI does not necessarily indicate an ability to evaluate its outputs critically. A sample of 17 participants was considered suitable for the qualitative objectives of the study because the emphasis was placed on obtaining detailed accounts of learners' reasoning processes rather than achieving statistical representativeness. The relatively small sample allowed close examination of individual responses while also permitting comparison across participants to identify recurring patterns, contrasting strategies, and atypical cases. The sample size therefore supported the study's aim of generating an in-depth account of critical reasoning within a clearly bounded instructional context.

Participant demographic information was collected to provide contextual interpretation of the findings. Relevant information included age, gender, educational level, length of English-language study, previous experience with online learning, frequency of academic writing activities, and prior experience with generative AI tools such as Gemini or ChatGPT. These characteristics were treated as contextual variables rather than as independent variables. They were used to describe the participant group and, where analytically relevant, to consider whether differences in previous digital, writing, or AI experience appeared to influence patterns

of critical engagement. No demographic characteristic was used to make assumptions about participants' critical-reasoning ability. Participation was voluntary, and learners were informed of the purpose of the research, the procedures involved, and their right to withdraw without academic consequences.

Instruments and Data Collection Technique

Two primary instruments were used to generate qualitative data: semi-structured interviews and classroom observations. The use of two complementary instruments allowed the study to examine both learners' reported reasoning and their observable engagement with Gemini during genre-based writing activities. A semi-structured interview protocol was developed in accordance with the research objectives and research questions. The interview questions focused on participants' experiences of evaluating Gemini-generated content during academic writing rather than merely their general attitudes toward artificial intelligence. The protocol was organized around several dimensions of critical reasoning: evaluation of factual accuracy, assessment of claim credibility, examination of evidence, identification of unsupported assumptions, recognition of inconsistencies, consideration of alternative arguments, evaluation of genre appropriateness, and justification of decisions to accept, reject, verify, or revise AI-generated information.

Questions also explored how the stages and instructional features of genre-based learning influenced learners' evaluation practices. Participants were therefore asked to reflect on how genre models, teacher explanations, collaborative discussion, rhetorical structures, evidence requirements, and revision activities affected their interactions with Gemini. Follow-up questions and probes were used to obtain clarification and concrete examples. For example, participants could be asked to explain why they trusted or questioned a particular Gemini response, how they verified a claim, what caused them to revise an AI-generated argument, and how they determined whether a response fitted the requirements of the academic genre.

The semi-structured format provided sufficient consistency across participants while allowing the interviewer to pursue unexpected but relevant issues that emerged from individual experiences. To strengthen content validity, the interview questions were mapped directly onto the research questions and the predetermined conceptual dimensions of critical reasoning. The wording of questions should also be reviewed for clarity, relevance, and neutrality by specialists familiar with EFL writing, qualitative research, or educational technology before formal data collection. Where feasible, a pilot interview with learners who share characteristics with the target participants can be used to identify ambiguous wording, repetitive questions, or prompts that unintentionally lead participants toward desirable answers. Any revisions resulting from expert review or piloting should be documented transparently.

Classroom observation was used to examine critical reasoning as it occurred during Gemini-assisted genre-based writing activities. An observation guide was developed from the same conceptual dimensions underpinning the interviews so that the two instruments generated complementary forms of evidence. The observer focused on instances in which learners questioned Gemini-generated information, compared AI responses with external evidence, challenged unsupported statements, identified inconsistencies, discussed alternative viewpoints, modified arguments, requested clarification from Gemini, or rejected content considered inappropriate for the target genre. The observation guide also documented how genre-based instructional activities appeared to stimulate these behaviors. Attention was given to learners' engagement with model texts, discussion of rhetorical moves, collaborative text construction, evidence evaluation, revision activities, and independent writing. Rather than recording only whether students used Gemini, the observations documented what students did with its outputs. A distinction was therefore maintained between passive adoption, superficial linguistic editing, information verification, argument evaluation, evidence-based revision, and independent reconstruction of AI-generated material.

Field notes were organized systematically according to lesson stage, observed action, relevant Gemini interaction, learner response, and preliminary analytical interpretation. Where ethically permitted, relevant classroom interactions could also be recorded to improve the accuracy of observational documentation. To enhance dependability, the observation categories were explicitly defined before data collection and used consistently across sessions. Interview and observation data were subsequently triangulated so that learners' reported critical practices could be examined against their observed behaviors. In qualitative research, validity and reliability were addressed through the related principles of credibility, dependability, confirmability, and consistency. Instrument alignment with the research questions strengthened content relevance, while expert review or piloting supported clarity and appropriateness. Triangulation across interviews and observations reduced dependence on a single source of evidence. Maintaining detailed field notes, coding definitions, analytical decisions, and revisions to the coding scheme provided an audit trail that increased transparency and dependability.

Data Analysis

Data from the semi-structured interviews and classroom observations were analyzed using an iterative qualitative process involving data condensation, data display, and conclusion drawing and verification, following the interactive analytical framework proposed by Miles, Huberman, and Saldaña. These processes were not treated as isolated sequential stages; rather, the researcher moved recursively between the data, emerging categories, displays, and preliminary interpretations throughout the analysis. Data condensation began with the organization and preparation of interview transcripts and observation field notes. Interview recordings were transcribed, anonymized, and read repeatedly together with observational records to develop familiarity with the dataset. Initial coding focused on segments that described or demonstrated learners' interaction with Gemini-generated information and arguments. Both deductive and inductive coding were employed. Deductive codes were informed by the research questions and conceptual dimensions of critical reasoning, including credibility evaluation, evidence checking, identification of unsupported claims, assumption recognition, logical evaluation, comparison of perspectives, verification, acceptance, rejection, revision, and genre appropriateness. Inductive codes were created when participants demonstrated reasoning behaviors or experiences not adequately represented by the initial framework.

The coding process also differentiated levels or forms of engagement. For example, accepting Gemini-generated content without examination was distinguished from modifying language only, checking information against another source, questioning evidentiary adequacy, restructuring an argument, or rejecting an AI response after critical examination. Related codes were subsequently clustered into broader categories and themes that represented recurring critical-reasoning practices. Following condensation, the coded data were organized into matrices and thematic displays to facilitate comparison within and across participants. One matrix could display participants in rows and critical-reasoning categories in columns, such as verification, evidence evaluation, identification of assumptions, counterargument consideration, genre alignment, and revision decisions. A second display could organize evidence according to stages of genre-based learning, allowing the researcher to examine where particular forms of critical reasoning occurred.

Interview statements and observational evidence were compared within these displays. For example, a participant who reported routinely verifying Gemini-generated information could be compared with observation records showing whether verification actually occurred during writing activities. Such comparison helped identify convergence, divergence, and exceptional cases and prevented the analysis from relying solely on self-reported perceptions. The final analytical process involved identifying broader patterns that addressed the two

research questions. For the first research question, interpretation focused on the strategies learners used to evaluate Gemini-generated information and arguments and the reasoning underlying decisions to accept, reject, verify, or revise particular outputs. For the second research question, analysis examined how specific genre-based learning activities were associated with critical engagement, including the use of genre models, collaborative discussion, rhetorical analysis, evidence requirements, and revision practices.

Preliminary conclusions were continuously checked against the complete dataset. Contradictory or negative cases were retained rather than removed because they provided important evidence about circumstances in which learners did not critically interrogate AI-generated content. Triangulation between interviews and observations was used to verify interpretations, while an audit trail documented coding decisions and changes to thematic categories. Where practicable, peer review of selected coded excerpts or member checking of interpretive summaries could further strengthen credibility. The final themes were selected on the basis of their relevance to the research questions, recurrence across participants or learning episodes, explanatory significance, and support from multiple forms of evidence. Direct participant quotations and representative observational episodes were then used to illustrate the themes. Through this process, the analysis moved beyond describing whether learners used Gemini toward explaining how they reasoned about its outputs, what prompted critical or uncritical engagement, and how genre-based learning shaped the evaluation and reconstruction of AI-generated content.

RESEARCH FINDINGS AND DISCUSSION

Research Findings

The qualitative analysis of interview and observational data was organized according to the two research questions. Following the procedures of data condensation, data display, and conclusion drawing, recurring patterns of learners' engagement with Gemini were grouped into themes representing different dimensions of critical reasoning. For RQ1, the analysis focused on how learners assessed the credibility, evidentiary quality, logical coherence, and genre appropriateness of Gemini-generated information. For RQ2, the analysis examined how genre-based learning created instructional conditions that encouraged learners to evaluate, select, revise, and integrate AI-generated material rather than accept it automatically.

RQ1: How do EFL learners critically evaluate the information and arguments generated by Gemini during genre-based academic writing tasks?

Evaluating the credibility and accuracy of Gemini-generated information

The first recurring pattern concerned learners' attempts to determine whether information generated by Gemini was sufficiently credible to be incorporated into academic writing. Learners did not consistently treat Gemini as an authoritative source. Instead, many appeared to regard its responses as preliminary information requiring additional verification. Their critical evaluation was particularly visible when Gemini generated factual claims, statistics, definitions, or references that seemed overly general or insufficiently substantiated. Learners described checking AI-generated information against course materials, scholarly articles, search results, or sources recommended by the teacher. This behavior demonstrated an emerging distinction between using Gemini as an idea-generating tool and considering it a reliable academic source. Observations also suggested that verification became more frequent when students were explicitly required to provide evidence for their claims. The following illustrative excerpts demonstrate this tendency:

P3: *"Gemini gives an answer very quickly, but I do not immediately put it in my paragraph. Sometimes I search the same information in an article because I need to know if the claim is really correct."*

P7: *"When Gemini gives a number or research result without a clear source, I become suspicious. I usually ask where the information comes from and then check it again."*
P11: *"I learned that a confident answer is not always an accurate answer. The language can sound very academic, but I still need evidence before I use it."*
P14: *"For general explanation, I can use Gemini to understand the topic, but for important facts I prefer to compare the answer with journal articles."*
P16: *"Sometimes the reference suggested by Gemini is difficult to find. If I cannot verify it, I remove that information from my writing."*

These responses suggest that critical evaluation involved an epistemic judgment about the status of AI-generated information. Rather than equating fluency with credibility, learners increasingly distinguished between plausible language and verifiable knowledge. The strongest critical behavior occurred when learners actively searched for corroborating evidence and rejected information that could not be substantiated.

Evaluating evidence and the strength of AI-generated arguments

A second theme concerned the distinction between a well-written argument and a well-supported argument. Learners reported that Gemini frequently generated claims that appeared logically persuasive at first glance but lacked adequate supporting evidence. During argumentative writing activities, participants increasingly examined whether examples, explanations, and evidence actually supported the claim being advanced. Some learners also recognized that Gemini tended to formulate broad assertions without clarifying the conditions or limitations under which those assertions were valid. Consequently, students began to ask whether evidence was relevant, sufficient, and appropriately connected to the claim. Representative illustrative responses included:

P1: *"At first I thought the argument was good because the paragraph sounded strong. But when I checked it, Gemini only repeated the idea and did not really give evidence."*
P5: *"I look at the claim first and then ask what evidence supports it. If the evidence is only another general statement, I do not consider the argument strong."*
P8: *"Sometimes Gemini gives examples, but an example is not always evidence. For academic writing, I need research or a source that can support the argument."*
P12: *"I noticed that Gemini can make a conclusion too quickly. I try to see whether the evidence really leads to that conclusion."*
P17: *"I ask Gemini for the opposing argument because if I only use one side, my paragraph becomes too simple. Seeing another perspective helps me evaluate the first argument."*

The findings indicate that learners' critical reasoning went beyond fact-checking and extended to argument evaluation. Learners began to inspect relationships between claims and evidence, distinguish illustration from empirical support, question overgeneralized conclusions, and consider alternative positions. This represents a shift from evaluating whether information is "correct" toward evaluating whether an argument is academically defensible.

Identifying inconsistencies, assumptions, and overly generalized claims

A third pattern involved learners' ability to identify weaknesses that were less immediately visible than factual inaccuracies. Several participants reported noticing contradictions between different Gemini responses, assumptions embedded in generated arguments, or generalizations that ignored contextual differences. Some students responded by reformulating prompts, requesting clarification, or comparing multiple AI-generated responses before deciding which information was usable. Observation data would be expected to show that more critical learners did not merely request a regenerated response but articulated why the original output was problematic. Illustrative interview excerpts included:

P2: *"When I asked the question in another way, Gemini gave a different answer. That made me realize that I should not depend on only one response."*
P6: *"Sometimes it says 'students usually benefit from AI,' but that is too general. I ask, 'Which students, in what situation, and according to what study?'"*
P9: *"I found that one paragraph said AI improves critical thinking, but another answer warned that AI can reduce thinking. I needed to understand the conditions instead of choosing one statement."*
P13: *"Gemini sometimes assumes something without explaining it. I try to identify what must be true for the argument to make sense."*
P15: *"If the statement uses words such as 'always' or 'all students,' I check it carefully because academic writing normally needs more cautious language."*

These examples illustrate a more analytical form of critical reasoning in which learners interrogated the conditions underlying a claim. They recognized that contradictions did not necessarily indicate that one answer was simply wrong; instead, differences could result from context, prompting, assumptions, or insufficient qualification. This encouraged learners to use more cautious and academically defensible language in their own writing.

Evaluating genre appropriateness and deciding what to accept, reject, or revise

The fourth theme demonstrated that learners evaluated Gemini-generated material not only for truthfulness but also for its suitability for the required academic genre. During genre-based writing activities, learners examined whether AI-generated content fulfilled the communicative purpose of a paragraph, followed an appropriate rhetorical sequence, maintained an academic tone, and included the expected components of argumentation. Students reported rejecting responses that were factually acceptable but rhetorically unsuitable. They also frequently revised AI-generated passages rather than incorporating them verbatim. Illustrative excerpts included:

P4: *"The information was relevant, but Gemini gave it like a general explanation. For my argumentative paragraph, I needed a clear claim and supporting evidence, so I changed the structure."*
P7: *"Sometimes the paragraph is too descriptive. I need to make it argumentative because the purpose of the assignment is to defend a position."*
P10: *"I compare the Gemini paragraph with the model text. If the organization is different from the genre we are learning, I reorganize it."*
P14: *"I do not keep every sentence. I select the useful idea, remove unnecessary explanation, and rewrite it in my own argument."*
P17: *"I ask whether the content helps the purpose of my paragraph. Even if the English is good, I delete it when it does not support my main point."*

This theme indicates that critical evaluation was inseparable from genre awareness. Learners evaluated AI output according to rhetorical function rather than linguistic quality alone. Their decisions to accept, reject, or revise Gemini-generated content therefore reflected both epistemic evaluation and genre-specific judgment.

Table 1. Results of EFL learners critically evaluation of the Information and Arguments generated by Gemini during Genre-Based Academic Writing Tasks

Theme	Critical-reasoning behavior	Typical learner action	Evidence expected from interviews/observations
Credibility and accuracy evaluation	Questioning reliability of AI-generated information	Checking claims against articles, course materials, or credible sources	Verification, source checking, rejection of unverifiable claims

Theme	Critical-reasoning behavior	Typical learner action	Evidence expected from interviews/observations
Evaluation of evidence and arguments	Examining whether evidence sufficiently supports claims	Distinguishing examples from evidence and assessing claim–evidence relationships	Argument checking, evidence comparison, alternative viewpoints
Identification of assumptions and inconsistencies	Detecting contradictions, overgeneralizations, and hidden assumptions	Re-prompting Gemini, comparing responses, qualifying claims	Recognition of inconsistency, contextualization, cautious language
Genre appropriateness	Assessing whether AI content fits the communicative purpose and rhetorical structure	Selecting, restructuring, rewriting, or deleting AI-generated material	Genre comparison, rhetorical revision, purposeful integration

The findings summarized in Table 1 suggest that learners' critical evaluation of Gemini-generated content operated at several interconnected levels. At the most immediate level, learners assessed factual credibility by verifying information through other sources. At a more advanced level, they evaluated the internal quality of arguments by examining relationships between claims and evidence, identifying generalizations and assumptions, and considering alternative interpretations. These practices indicate that critical reasoning cannot be reduced to detecting false information; it also involves determining whether knowledge claims are sufficiently supported and logically defensible.

A particularly important finding is the role of genre awareness in shaping learners' final decisions about AI-generated content. Information could be factually acceptable but still be rejected because it did not fulfill the rhetorical purpose of the assignment. Thus, the evaluation process involved at least two related judgments: "Is this information credible?" and "Does this information function appropriately in this genre?" The interaction of these judgments suggests that critical AI literacy in academic writing should incorporate epistemic, argumentative, and rhetorical dimensions rather than focus exclusively on factual verification.

RQ2: How does genre-based learning support EFL learners' critical reasoning when they evaluate, select, revise, and integrate Gemini-generated content into their academic writing?

Building knowledge of the field provided a basis for questioning Gemini

The first theme showed that the building-knowledge stage of genre-based learning provided learners with background knowledge necessary to evaluate Gemini-generated responses. Students indicated that it was difficult to recognize inaccurate, incomplete, or oversimplified AI output when they had little prior knowledge of the topic. Exposure to readings, key concepts, vocabulary, and source materials before interacting extensively with Gemini gave learners a reference point for judging AI-generated content. Instead of using Gemini to replace topic learning, students became better positioned to compare generated information against what they had already learned. Illustrative excerpts included:

P1: *"Before reading about the topic, almost every Gemini answer looked correct to me. After reading the articles, I could see which parts were too general."*
P5: *"The background materials helped me ask better questions. I knew what information was missing from Gemini's first answer."*
P9: *"When I already understood the important concepts, I did not just follow Gemini. I could disagree with some parts and explain why."*

P13: *"Learning the topic first made Gemini more useful because I could compare its answer with what we studied instead of depending on it."*
P16: *"I think knowledge is necessary for critical thinking. If I know nothing about the topic, I cannot know whether Gemini gives me a weak answer."*

These responses illustrate that critical evaluation depended partly on domain knowledge. Genre-based learning supported critical reasoning by ensuring that learners possessed sufficient conceptual resources to interrogate AI output. Gemini therefore became a supplementary resource rather than the primary authority defining the content of the writing task.

Analysis of model texts provided explicit criteria for evaluating AI-generated writing

The second theme concerned the role of model texts. Genre-based instruction exposed learners to examples of successful academic writing and made rhetorical expectations visible. Participants reported comparing Gemini-generated paragraphs with these models to determine whether claims were clearly presented, evidence was appropriately positioned, counterarguments were addressed, and conclusions followed logically. Model analysis therefore transformed genre conventions into practical criteria for evaluating AI-generated writing. Illustrative excerpts included:

P3: *"The model text showed me that an argument needs more than a topic sentence. When Gemini gave only a claim and explanation, I knew something was missing."*
P6: *"I used the model to check the organization. I asked whether Gemini had a claim, evidence, and explanation in the correct order."*
P8: *"Before learning the genre, I mostly checked grammar. After analyzing the model texts, I started checking what each sentence was doing in the paragraph."*
P11: *"The example text helped me see that academic arguments also need qualification and sometimes a counterargument. Gemini's first response was too certain."*
P15: *"I compared the AI paragraph with the genre structure from class. That made it easier to decide which parts should be moved, removed, or expanded."*

The findings suggest that model texts functioned as evaluative benchmarks. Learners shifted from judging AI output impressionistically—whether it "sounded good"—toward judging it according to explicit rhetorical criteria. Genre knowledge consequently made critical reasoning more systematic and less dependent on intuitive perceptions of language quality.

Joint construction and classroom discussion encouraged justification and alternative perspectives

The third theme showed that collaborative stages of genre-based learning supported critical reasoning through dialogue. During joint construction, learners discussed whether Gemini-generated claims should be retained, modified, supported with additional evidence, or rejected. Such interaction required learners to verbalize and defend their judgments. Peer disagreement was particularly valuable because it exposed participants to alternative interpretations and reduced the likelihood that the first plausible AI response would be accepted automatically. Illustrative excerpts included:

P2: *"When I worked alone, I sometimes accepted Gemini's suggestion. In the group, another student asked where the evidence was, so we checked it before using it."*
P4: *"Discussion made me explain why I thought one argument was stronger. I could not just say, 'Gemini gave it.' I needed a reason."*
P7: *"My friend interpreted the same answer differently. We compared both ideas and then changed the paragraph to make the argument clearer."*

P12: *"Sometimes we disagreed with Gemini and also with each other. That discussion helped us see different sides of the issue."*
P17: *"The teacher often asked, 'Why do you accept this sentence?' That question made us think about evidence, not only grammar."*

These responses suggest that collaborative construction externalized the reasoning process. Critical judgments that might otherwise remain implicit became subjects of negotiation. Teacher questioning and peer dialogue therefore functioned as scaffolds that encouraged justification, comparison, and reconsideration of AI-generated content.

Independent construction encouraged selective integration and learner ownership

A fourth pattern emerged during independent writing. After receiving earlier scaffolding, learners increasingly described Gemini as a resource from which they selected and transformed information rather than as a producer of final text. Independent construction required them to make individual decisions about what content to retain, verify, reorganize, paraphrase, or reject. Several participants suggested that revising Gemini output helped them maintain ownership of the argument and align the final text with their intended position. Illustrative excerpts included:

P4: *"In the final writing, I used Gemini for some ideas, but I did not use the whole response. I chose only what supported my argument."*
P6: *"I changed many sentences because I wanted the paragraph to express my position, not Gemini's position."*
P10: *"Sometimes I completely rejected the AI suggestion because after checking the evidence, I found another argument that was stronger."*
P13: *"I combined information from the reading, my own idea, and some suggestions from Gemini. The final paragraph was not taken from one source."*
P15: *"Independent writing made me responsible for the final decision. Gemini could recommend something, but I had to decide if it belonged in the text."*

This pattern indicates that genre-based scaffolding could shift human–AI interaction from passive adoption toward selective integration. Learners retained responsibility for rhetorical and epistemic decisions while using Gemini as one resource among several. The result was not necessarily reduced AI use but a more deliberate form of AI use.

Reflection and revision strengthened metacognitive awareness of AI use

The final theme involved reflection on the process of interacting with Gemini. Revising drafts and explaining changes prompted learners to examine not only the quality of their texts but also their own decision-making. Several participants became more aware of when they depended too heavily on Gemini and when AI assistance genuinely supported learning. Reflection encouraged students to articulate criteria for responsible AI use and recognize situations in which independent reasoning should take priority. Illustrative excerpts included:

P3: *"When I looked at my first draft, I realized that too many ideas came directly from Gemini. In revision, I checked which ideas I could really explain myself."*
P8: *"The reflection made me think about why I accepted some answers quickly. Sometimes it was because the English sounded professional."*
P11: *"I learned that using AI critically takes more time than copying, because I have to check, compare, and revise."*
P14: *"Now I ask myself whether Gemini is helping me think or doing the thinking for me."*
P16: *"After revision, I became more careful. I use Gemini to support my writing, but I try to keep control of the argument."*

These responses suggest that metacognitive reflection reinforced the distinction between assistance and substitution. Learners increasingly evaluated not only Gemini's responses but

also their own patterns of reliance on AI. Genre-based revision thus provided a context for developing awareness of human responsibility within AI-assisted writing.

Table 2. Results of EFL Learners' Critical Reasoning Integrated with Gemini-Generated Content into Academic Writing

Genre-based learning component	Contribution to critical reasoning	Typical learner behavior	Resulting function in Gemini-assisted writing
Building knowledge of the field	Provides conceptual knowledge for evaluating AI claims	Comparing Gemini output with readings and prior knowledge	Detection of inaccuracies, omissions, and oversimplifications
Analysis of model texts	Provides explicit rhetorical and genre criteria	Comparing AI paragraphs with model structures	Evaluation of organization, evidence placement, and genre fit
Joint construction	Promotes dialogue, justification, and consideration of alternatives	Discussing, questioning, and defending decisions about AI output	Collaborative verification and argument refinement
Independent construction	Transfers responsibility for evaluative decisions to learners	Selecting, rejecting, revising, and synthesizing AI-generated ideas	Greater learner ownership and selective AI integration
Reflection and revision	Develops metacognitive awareness of AI dependence and decision-making	Reviewing why AI suggestions were accepted or rejected	More deliberate and responsible use of Gemini

Table 2 indicates that genre-based learning supported critical reasoning not through a single instructional activity but through a sequence of complementary scaffolds. Building knowledge provided the conceptual basis for questioning AI, while model-text analysis supplied rhetorical criteria for evaluating the quality of generated arguments. Joint construction added a social dimension by requiring learners to justify decisions and respond to competing interpretations. Independent construction subsequently transferred responsibility from teachers and peers to individual learners, requiring them to determine how AI-generated material would contribute to their own argument.

The findings further suggest that the contribution of genre-based learning lies in transforming AI use from a predominantly production-oriented activity into an evaluation-oriented process. Gemini was not simply used to increase the speed or linguistic sophistication of writing. Instead, the pedagogical stages repeatedly required learners to interrogate information, compare alternatives, justify rhetorical choices, and reflect on their dependence on AI. The gradual movement from supported analysis to independent construction appears particularly important because it maintains the benefits of AI assistance while preserving learner responsibility for evidence, reasoning, and textual decisions.

Discussion

The findings demonstrate that EFL learners' critical engagement with Gemini extended considerably beyond checking whether AI-generated information was factually correct. Participants evaluated credibility, examined relationships between claims and evidence, questioned assumptions and overgeneralizations, and assessed whether generated material fulfilled the rhetorical requirements of the target academic genre. The manuscript therefore presents critical reasoning as an interconnected epistemic, argumentative, and rhetorical process. This interpretation is consistent with Jiang's (2026) conceptualization of GenAI literacy in EFL writing as involving evaluation rather than merely functional competence with

AI tools. It also supports broader arguments that effective AI literacy requires learners to understand both the capabilities and limitations of generative systems and to exercise human judgment over generated content (Adamakis & Rachiotis, 2025). The present findings extend this position by showing how such judgment may become observable in concrete writing decisions: verifying information, rejecting unsupported claims, qualifying overgeneralizations, restructuring arguments, and determining whether AI-generated content serves the purpose of a particular genre. This is especially important because the participants did not consistently equate Gemini's linguistic confidence with epistemic reliability; instead, some learners explicitly distinguished polished academic language from verifiable knowledge.

The first major finding, therefore, concerns the development of source and credibility evaluation. Learners reported checking Gemini-generated claims against scholarly articles, course materials, and other credible sources, particularly when AI responses included factual claims, statistics, or references that could not immediately be substantiated. This finding confirms concerns raised in previous research that GenAI can produce fluent responses whose apparent authority may conceal inaccuracies, unsupported information, or unreliable references (Faisal, 2026; Han, 2025; Tang, 2024). It also reinforces the distinction identified by Ma and Vasinda (2025) between the efficiency afforded by GenAI and the need for pedagogical structures that prevent students from accepting generated material uncritically. However, the present study adds a useful learner-centered dimension to this literature. Instead of examining misinformation only as a technological limitation, the findings show how uncertainty in AI output can itself become a trigger for critical inquiry when students are expected to substantiate claims. Thus, the educational problem is not simply that Gemini may generate unreliable information; the more important issue is whether instructional conditions require learners to recognize, investigate, and resolve that uncertainty.

A second important finding was that learners distinguished between linguistically persuasive arguments and evidentially defensible arguments. Participants identified occasions when Gemini generated paragraphs that appeared convincing but merely repeated claims, relied on examples rather than evidence, or moved too quickly from information to conclusions. This finding is consistent with Hao et al. (2026), who showed that higher-order thinking in AI-supported argumentative writing depends on instructional arrangements that go beyond automated feedback, and with Tang (2024), who argued that generative AI does not automatically produce deeper analytical engagement. The result also complements Li's (2024) work on AI-supported L2 writing feedback by indicating that linguistic improvement and argument quality are analytically distinct outcomes. Gemini may assist with cohesion, grammatical accuracy, and organization, but these affordances do not guarantee that evidence adequately supports a claim. The present study therefore extends existing scholarship by emphasizing the claim-evidence relationship as a central site of human judgment. Critical use of GenAI appears to require learners to ask not only whether the AI response is relevant but also whether the warrant connecting evidence and conclusion is academically defensible.

The participants' attention to assumptions, contradictions, and overgeneralization further indicates a more advanced form of critical engagement. Learners noticed that differently phrased prompts could elicit different answers and that apparently contradictory AI claims sometimes reflected changes in context, scope, or unstated assumptions. This finding adds nuance to concerns about hallucination and inconsistency in GenAI research. Rather than treating inconsistency only as a defect to be eliminated, learners could use conflicting outputs as an opportunity to interrogate conditions under which a claim might or might not hold. Such behavior aligns with Jiang's (2026) emphasis on evaluative AI literacy and with Yang et al.'s (2025) argument that GenAI-assisted L2 writing should be examined in relation to higher-order processes, not only performance outcomes. The participants' tendency to replace categorical expressions such as "always" or "all students" with more qualified academic language also

suggests that critical reasoning and genre awareness were mutually reinforcing. Epistemic caution became visible rhetorically through hedging, qualification, and contextualization.

The findings for RQ1 also show that critical judgment was shaped by genre expectations. Participants sometimes rejected content that was factually plausible because it did not perform the rhetorical function required by an argumentative text. They compared Gemini-generated passages with model texts, reorganized descriptive responses into argumentative structures, deleted irrelevant information, and retained only material that supported their central position. This finding supports Zhang and Samsudin's (2026) argument that AI literacy can be productively integrated with genre-based pedagogy. It also extends that work by shifting attention from teacher AI literacy to learners' actual evaluative decisions. Genre knowledge appears to provide a practical framework for deciding what to do with AI output. Consequently, the question "Is this accurate?" was complemented by "Does this content function appropriately here?" This dual evaluation is theoretically important because it suggests that critical AI literacy in academic writing includes not only epistemic validity but rhetorical fitness.

The findings for RQ2 further clarify why genre-based learning may facilitate this form of reasoning. The building-knowledge stage supplied learners with conceptual resources against which Gemini responses could be evaluated. Participants explicitly indicated that they found it difficult to challenge AI-generated information when they lacked sufficient background knowledge, whereas prior engagement with readings and key concepts enabled them to recognize oversimplification and missing information. This supports the broader pedagogical principle that critical thinking is partly knowledge-dependent rather than a completely generic skill. It is also consistent with research emphasizing the need to integrate GenAI with disciplinary knowledge instead of treating prompting ability as an adequate form of AI literacy (Chen, 2025; Jiang, 2026). A significant implication is that GenAI should not replace the knowledge-building phase of genre pedagogy. If learners approach AI without a sufficient conceptual base, they may lack the resources necessary to evaluate what the system produces.

Analysis of model texts provided a second layer of scaffolding by converting genre expectations into explicit evaluative criteria. Participants reported shifting from asking whether Gemini-generated text "sounded good" to examining what individual sentences were doing rhetorically, whether evidence was positioned appropriately, and whether claims, qualifications, and counterarguments were adequately represented. This finding strengthens the case for integrating GenAI into established writing pedagogy rather than developing separate "AI activities." Model-text analysis gave learners a benchmark against which AI output could be assessed, thereby reducing the authority of fluency as the dominant criterion of quality. This finding is consistent with Divya and Alamelu (2026), who reported pedagogical value in integrating AI tools into argumentative writing, but the current study adds that the benefit is likely to depend on whether learners have explicit rhetorical criteria for judging AI-generated content.

Joint construction and classroom discussion constituted another important mechanism. Learners were required to explain why they accepted one argument, rejected another, or requested additional evidence. Peer disagreement and teacher questioning made reasoning visible and negotiable. This result is compatible with Liu et al. (2026), whose work on GenAI-supported collaborative composing highlights the importance of interaction in AI-mediated learning. It also supports Mi et al.'s (2025) emphasis on the complementary value of human and AI feedback. In the present study, peers and teachers did not merely provide alternative answers; they created accountability for reasoning. A learner could no longer justify a claim simply by saying that Gemini generated it. The requirement to articulate a rationale changed the status of AI output from answer to proposition—a proposition that had to survive scrutiny.

The progression toward independent construction and reflection further suggests that effective AI integration does not necessarily require reducing students' access to AI. Rather, it requires transferring evaluative responsibility to learners. Participants described selecting useful ideas, rejecting weak suggestions, combining AI-generated material with scholarly sources and their own reasoning, and revising passages to maintain control over the argument. This finding aligns with Chen and Alibakhshi's (2025) emphasis on cognitive and metacognitive strategies in AI-supported language learning and with Roe et al.'s (2026) argument for graduated frameworks of AI use in academic contexts. Importantly, the outcome was not complete independence from Gemini but more deliberate interdependence: learners continued to use AI while assuming responsibility for the final epistemic and rhetorical decisions. This challenges simplistic binaries between AI dependence and AI prohibition.

Taken together, the findings suggest that the principal pedagogical contribution of genre-based learning is its capacity to transform GenAI use from a production-oriented practice into an evaluation-oriented practice. The sequence of building knowledge, analyzing models, jointly constructing arguments, independently composing texts, and reflecting on revision repeatedly positioned learners as evaluators rather than recipients of AI output. This extends previous work that has primarily examined GenAI in relation to writing quality, learner perception, feedback, or technology acceptance (Li, 2024; Mi et al., 2025; Wang et al., 2025; Xiao & Zhu, 2025). The contribution of the present study lies in showing how genre-based scaffolding can organize human–AI interaction around critical reasoning. The findings therefore support a model in which Gemini is neither treated as an authoritative writer nor rejected as pedagogically undesirable. Instead, it becomes a deliberately imperfect interlocutor whose outputs provide material for verification, argument testing, rhetorical evaluation, revision, and metacognitive reflection.

Nevertheless, the findings should be interpreted within the boundaries of the qualitative case-study design. The study involved 17 EFL learners in Iraqi private-school settings, selected purposively because of their experience with online learning and academic writing; consequently, the findings offer contextualized analytical insights rather than statistical generalization. Learner proficiency, prior AI experience, disciplinary knowledge, teacher mediation, and genre could produce different patterns in other contexts. Future studies could compare proficiency levels, genres, or AI platforms and examine whether the observed evaluative practices persist longitudinally after instructional scaffolding is reduced. Nevertheless, the present findings indicate that the most educationally productive question may no longer be whether EFL learners should use GenAI, but under what pedagogical conditions its use requires them to reason more carefully rather than think less.

CONCLUSION

This study investigated how EFL learners critically evaluated information and arguments generated by Gemini during genre-based academic writing and how genre-based learning supported those evaluative processes. The findings indicate that critical engagement with Gemini operated across several interconnected dimensions. Learners evaluated the credibility and accuracy of generated information, examined the adequacy of evidence and claim-evidence relationships, identified assumptions, contradictions, and overgeneralizations, and judged whether generated material fulfilled the communicative and rhetorical requirements of the target genre. These practices demonstrate that critical reasoning in GenAI-assisted writing extends beyond detecting incorrect information. It involves decisions about whether knowledge is credible, whether arguments are defensible, whether evidence is sufficient, and whether particular content functions appropriately within an academic genre. Gemini therefore became most educationally useful when its outputs were treated as provisional propositions requiring human verification and judgment rather than as authoritative text ready for

incorporation into students' writing. The study consequently reinforces the need to distinguish the linguistic fluency of AI-generated text from its epistemic and argumentative quality.

Genre-based learning appeared to provide the instructional structure necessary for developing this evaluative orientation. Building knowledge of the field equipped learners with conceptual resources for questioning AI-generated claims; model-text analysis supplied explicit rhetorical criteria; joint construction promoted justification and consideration of alternatives; independent construction transferred responsibility for selection and revision to learners; and reflection strengthened awareness of the difference between productive AI assistance and cognitive substitution. The study therefore contributes to emerging research on GenAI-assisted EFL writing by linking Gemini use, critical reasoning, and genre pedagogy within a single pedagogical framework. Its practical implication is that educators should move beyond teaching learners how to prompt GenAI or improve the language of generated texts. Academic writing tasks should systematically require verification, evidence appraisal, argument evaluation, rhetorical justification, and reflective decision-making. Under such conditions, GenAI can function not as a replacement for reasoning but as material through which reasoning is exercised and made visible. Because the study was context-specific and qualitative, future research should investigate these processes across different proficiency levels, genres, institutional settings, and AI platforms and examine whether critical practices remain stable over time.

ACKNOWLEDGEMENT

The authors express their appreciation to the EFL learners who participated in the study and contributed their experiences and perspectives through the interviews and classroom activities. Their participation made it possible to examine how learners engage critically with Gemini-generated information and arguments in genre-based academic writing. The authors also acknowledge the academic and institutional support that facilitated the conduct of the study and the completion of the manuscript. Any specific funding agency, school, research center, or institutional permission should be named here only if such support was actually provided.

CONFILICT INTEREST

There is no conflict of interest

RESEARCH FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. All research activities, including data collection, analysis, preparation of the manuscript, and publication-related processes, were supported independently by the authors. The funders had no role in the conduct of this study.

INFORMED CONSENT STATEMENT

Participation in this study was entirely voluntary. Prior to participation, all participants were informed about the purpose, procedures, potential risks, and benefits of the study. Participants were assured that their identities and personal information would remain confidential and that all information provided would be used solely for research purposes.

DATA AVAILABILITY STATEMENT

The data generated and analyzed during this study are not publicly available due to privacy, confidentiality, and ethical considerations related to the protection of research participants. Access to the dataset may be provided by the corresponding author upon reasonable request for purposes such as verification or further scholarly analysis. Each request will be evaluated on a case-by-case basis, and data will only be shared following approval from the appropriate institutional ethics review board and in accordance with applicable ethical requirements, data protection regulations, and the terms of participant consent.

DECLARATION OF USING AI TOOLS

Generative artificial intelligence tools were used to assist with language refinement, organization, and the drafting and revision of selected sections of this manuscript. All AI-assisted content was reviewed, verified, and revised by the authors, who remain fully responsible for the accuracy, interpretation, originality, and integrity of the final article. AI tools were not treated as scholarly evidence or substitutes for authorial responsibility.

REFERENCES

Adamakis, M., & Rachiotis, T. (2025). Artificial intelligence in higher education: A state-of-the-art overview of pedagogical integrity, artificial intelligence literacy, and policy integration. *Encyclopedia, 5*(4), 180. https://doi.org/10.3390/encyclopedia5040180

Barrett, A., & Pack, A. (2023). Not quite eye to A.I.: Student and teacher perspectives on the use of generative artificial intelligence in the writing process. *International Journal of Educational Technology in Higher Education, 20*(1). https://doi.org/10.1186/s41239-023-00427-0

Chen, F. (2025). Scientific mapping of GenAI in English as a foreign language (EFL) context. *International Journal of Technology and Human Interaction, 21*(1), 1–23. https://doi.org/10.4018/IJTHI.384377

Chen, J., & Alibakhshi, G. (2025). AI-powered applications' effects on English language learners' cognitive, metacognitive, and resource management strategies, and language achievement. *Journal of Computer Assisted Learning, 42*(1). https://doi.org/10.1002/jcal.70171

Divya, T. J., & Alamelu, C. (2026). Integrating DeepL Write and Claude AI to enhance argumentative writing competence of engineering students in India. *Frontiers in Artificial Intelligence, 9*. https://doi.org/10.3389/frai.2026.1838463

Faisal, E. (2026). Complementary AI in higher education: Behavioral, cognitive, and ethical implications of ChatGPT and DeepSeek. *Frontiers in Psychology, 17*. https://doi.org/10.3389/fpsyg.2026.1699114

Han, J. (2025). Are teachers assessing work written by students or by AI? A rapid literature review of research on detecting content generated by generative AI. *European Journal of Education, 60*(4). https://doi.org/10.1111/ejed.70240

Hao, H., Razali, A. B., & Zuo, R. (2026). Exploring the impact of integrated AWE and generative AI feedback on Chinese EFL undergraduates' higher-order thinking in argumentative writing. *SAGE Open, 16*(1). https://doi.org/10.1177/21582440251413884

Jiang, G. (2026). A conceptual analysis: Generative artificial intelligence literacy in English as a foreign language writing. *Frontiers in Psychology, 17*. https://doi.org/10.3389/fpsyg.2026.1815271

Kim, J., Yu, S., Detrick, R., & Li, N. (2024). Exploring students' perspectives on generative AI-assisted academic writing. *Education and Information Technologies, 30*(1), 1265–1300. https://doi.org/10.1007/s10639-024-12878-7

Lee, I. B. H., Mak, P., & Yu, S. (2026). Feedback in writing in the globalized era: Perspectives, opportunities and challenges. *RELC Journal, 57*(1), 3–12. https://doi.org/10.1177/00336882261423572

Li, M. (2024). Leveraging ChatGPT for second language writing feedback and assessment. *International Journal of Computer-Assisted Language Learning and Teaching, 14*(1), 1–11. https://doi.org/10.4018/IJCALLT.360382

Liu, M., Zhang, L. J., & Biebricher, C. (2026). Exploring ESL learners' peer interactions in GenAI-supported collaborative multimodal PowerPoint composing. *Written Communication, 43*(4), 1121–1158. https://doi.org/10.1177/07410883261461544

Lu, C. (2025). AI-generated corpus learning and EFL learners' learning of grammatical structures, lexical bundles, and willingness to write. *PLOS ONE, 20*(7), e0321544. https://doi.org/10.1371/journal.pone.0321544

Ma, C., & Vasinda, S. (2025). GenAI in academic writing: Empowering learners or redefining traditional pedagogical practices? *International Journal of Artificial Intelligence, 1*(1), 1–34. https://doi.org/10.4018/IJAITL.373582

Mi, Y., Rong, M., & Chen, X. (2025). Exploring the affordances and challenges of GenAI feedback in L2 writing instruction: A comparative analysis with peer feedback. *ECNU Review of Education, 9*(1). https://doi.org/10.1177/20965311241310883

Mohammadkarimi, E. (2023). Teachers' reflections on academic dishonesty in EFL students' writings in the era of artificial intelligence. *Journal of Applied Learning & Teaching, 6*(2), 105–113. https://doi.org/10.37074/jalt.2023.6.2.10

Roe, J., Perkins, M., & Tregubova, Y. (2026). The EAP-AIAS: Adapting the AI Assessment Scale for English for Academic Purposes. *TESOL Journal, 17*(2). https://doi.org/10.1002/tesj.70122

Slimi, Z., Alawai, F. A., Alyani, H. A., Abri, S. A., Al-Farsi, F. A., & Balushi, K. A. (2022). Writing issues in ESL and their potential solutions: Case study IMCO's foundation students. *Journal of Educational and Social Research, 12*(6), 81. https://doi.org/10.36941/jesr-2022-0146

Tang, K. (2024). Informing research on generative artificial intelligence from a language and literacy perspective: A meta-synthesis of studies in science education. *Science Education, 108*(5), 1329–1355. https://doi.org/10.1002/sce.21875

Tang, W., Wang, Z., & Guangfen, Y. (2026). What are we missing in AI-supported L2 writing research? A systematic review from 2014 to 2024. *Journal of Computer Assisted Learning, 42*(3). https://doi.org/10.1002/jcal.70262

Urazbayeva, G., Kussainova, R., Aibergen, A., Kaliyeva, A., & Kantayeva, G. (2024). Innovation off the bat: Bridging the ChatGPT gap in digital competence among English as a foreign language teachers. *Education Sciences, 14*(9), 946. https://doi.org/10.3390/educsci14090946

Wagner, G., Prester, J., Mousavi, R., Lukyanenko, R., & Paré, G. (2026). Generative artificial intelligence for literature reviews. *Journal of Information Technology, 41*(2), 148–170. https://doi.org/10.1177/02683962261425675

Wang, C., Li, X., & Zou, B. (2025). Revisiting integrated model of technology acceptance among the generative AI-powered foreign language speaking practice: Through the lens of positive psychology and intrinsic motivation. *European Journal of Education, 60*(1). https://doi.org/10.1111/ejed.70054

Xiao, F., & Zhu, S. (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, 9*(1). https://doi.org/10.1177/20965311241310881

Yang, C., Li, R., & Yang, L. (2025). Revisiting trends in GenAI-assisted second language writing: Retrospect and prospect. *Journal of Educational Computing Research, 63*(7–8), 1819–1863. https://doi.org/10.1177/07356331251367309

Zaimoğlu, S., & Dağtaş, A. (2025). Teacher cognition and practices in using generative AI tools to support student engagement in EFL higher-education contexts. *Behavioral Sciences, 15*(9), 1202. https://doi.org/10.3390/bs15091202

Zhang, Y., & Samsudin, M. A. (2026). From familiarity to criticality: Cultivating EFL teachers' AI literacy through an AI-integrated genre-based pedagogy. *Education Sciences, 16*(1), 150. https://doi.org/10.3390/educsci16010150