Ilmiy tadqiqotlar va ularning yechimlari jurnali
Aoka guvohnoma № 323759 ISSN: 3060-4818

INTEGRATING GENERATIVE ARTIFICIAL INTELLIGENCE INTO SECONDARY-SCHOOL ENGLISH LANGUAGE TEACHING: A METHODOLOGICAL FRAMEWORK FOR DEVELOPING CRITICAL THINKING AND AI LITERACY

Hazratova Aziza
English Language Teacher,
School No. 353, Tashkent, Uzbekistan
https://doi.org/10.5281/zenodo.23376781

Article type: methodological/conceptual article with a proposed, not yet validated, classroom intervention.

Important research-integrity note: This manuscript does not report a completed experiment. The intervention and assessment plan are proposed for future implementation. No student sample, results, statistical significance, or causal effect is invented. Before submission, replace this note only after conducting the study and reporting verified data.

Abstract

Generative artificial intelligence (GenAI) offers new possibilities for English language teaching, including text generation, language practice, and feedback. However, access to generated answers does not itself ensure that learners can judge their accuracy, relevance, or limitations. This article proposes a methodological framework for integrating GenAI into secondary-school English lessons to support critical thinking and AI literacy. The framework is organised as a five-stage learning cycle: Generate, Verify, Evaluate, Argue, and Reflect (GVEAR). Learners first elicit an AI-generated response, identify claims that require checking, compare the response with suitable evidence, justify their own position in English, and reflect on the tool's limitations and their own decisions. The article situates the framework within research on generative AI in education, language teaching, critical thinking, and AI literacy. It also specifies a proposed quasi-experimental design, participant safeguards, assessment dimensions, lesson sequence, and data-analysis plan that a school-based researcher could use to evaluate the framework. Because no intervention has yet been conducted for this manuscript, the framework is presented as a researchable proposal rather than as an empirically validated method. Its contribution is a transparent alignment of language-learning tasks with explicit critical-evaluation behaviours and responsible AI use.

Keywords: generative artificial intelligence; English language teaching; critical thinking; AI literacy; secondary education; teaching methodology; quasi-experimental design.

1. Introduction

Generative artificial intelligence (GenAI) systems can produce explanations, dialogues, summaries, examples, and feedback in response to natural-language prompts. In English language teaching, these affordances may help teachers prepare differentiated materials and give learners additional opportunities to practise reading, writing, vocabulary, and interaction. Kohnke, Moorhouse, and Zou (2023) discuss both the potential of ChatGPT for language teaching and the digital competencies required for its ethical and effective use. Kasneci et al. (2023) similarly identify educational opportunities alongside concerns such as unreliable output, bias, privacy, and over-reliance.

The educational question is therefore not simply whether learners should use GenAI, but what learning design can make such use purposeful. If a learner copies an AI-generated paragraph without examining its claims or language, the technology may complete a task without ensuring that the learner has developed the underlying competence. By contrast, when learners compare responses, verify claims, identify weaknesses, and justify revisions, AI output can become material for analysis rather than a substitute for thinking. UNESCO (2023) recommends a human-centred approach that protects learner agency and requires attention to pedagogical suitability, privacy, inclusion, and age-appropriate use.

This issue matters in secondary-school English classrooms because language learning offers authentic opportunities to practise evaluation and argumentation in the target language. Learners can assess whether a generated explanation is sufficiently supported, whether a claim is too general, whether a source is credible, and whether a proposed conclusion follows from the evidence. These activities can combine language objectives with higher-order cognitive objectives, provided that tasks are age-appropriate and teacher-guided.

The research problem addressed in this article is the methodological challenge of moving from occasional AI use to a structured sequence of learning activities in which English language practice and critical evaluation are explicitly connected. The article does not claim that this gap is absent from all previous scholarship; rather, it proposes a classroom framework that can be examined in a defined secondary-school context.

Aim. To develop a theoretically grounded methodological framework for integrating GenAI into secondary-school English lessons to support the development of critical thinking and AI literacy.

Research questions. (1) What stages can structure GenAI-supported English tasks so that learners evaluate rather than merely accept AI output? (2) Which observable indicators can be used to assess critical thinking and AI literacy during such tasks? (3) How could the proposed framework be evaluated in a school-based study?

The article makes a conceptual and methodological contribution: it specifies a five-stage learning cycle, aligns each stage with observable learner actions, and outlines a transparent design for future empirical evaluation. The framework's effectiveness remains an empirical question.

2. Theoretical Background and Literature Review

2.1. Generative AI in language education

GenAI can support language learning by producing model texts, alternative explanations, role-play prompts, vocabulary examples, and feedback. These affordances may increase the range of materials available to teachers and create additional opportunities for practice. Nevertheless, generated text can contain factual errors, inappropriate language, unsupported claims, or responses that do not fit a learner's context. Kohnke et al. (2023) therefore emphasise the need to consider both the opportunities and limitations of ChatGPT in language teaching. Kasneci et al. (2023) also argue that educational use of large language models must be considered alongside risks and responsible-use requirements.

For methodology, this means that AI should be linked to a defined learning objective and a task that requires learner action. A teacher might ask students to compare an AI-generated text with a reliable source, revise a paragraph for a specified audience, or explain why one answer is better supported than

another. The teacher remains responsible for selecting appropriate tasks, monitoring output, and ensuring that AI use does not replace the target learning activity.

2.2. Critical thinking as an assessable learning outcome

Critical thinking is not a single observable behaviour. Facione's (1990) Delphi report identifies core skills including interpretation, analysis, evaluation, inference, explanation, and self-regulation. For the present framework, critical thinking is operationalised through five classroom indicators: identifying claims, checking evidence, comparing alternatives, giving a reasoned justification, and revising a conclusion when evidence requires it. These indicators are not presented as a new universal definition of critical thinking; they are task-level indicators selected to make the proposed intervention assessable.

In English lessons, these indicators can be elicited through short reading texts, discussion prompts, evidence-based writing, and peer review. For example, students may receive an AI-generated paragraph containing one unsupported generalisation and one inaccurate claim. Their task is not merely to label the paragraph as wrong, but to identify the relevant statement, provide a reason for questioning it, and revise it in clear English.

2.3. AI literacy and responsible use

AI literacy includes understanding what AI systems can and cannot do, recognising that output may be inaccurate, and making informed decisions about when and how to use the technology. Long and Magerko (2020) conceptualise AI literacy as a set of competencies that enables people to critically evaluate AI technologies, communicate and collaborate with AI, and use AI as a tool. In a school setting, the concept should be translated into age-appropriate behaviours rather than taught only as abstract terminology.

For this article, AI literacy is operationalised through four dimensions: (1) understanding that generated content may be inaccurate; (2) writing a prompt that states a clear task and constraints; (3) verifying important claims with suitable evidence; and (4) using AI transparently and responsibly. These dimensions are a practical assessment proposal and should be reviewed by subject experts before use.

2.4. Rationale for an integrated framework

The theoretical rationale is that language learning, critical thinking, and AI literacy can be brought together when the learner must work on AI output in a purposeful English-language task. The framework proposed below combines generation with verification, evaluation, argumentation, and reflection. Its value should not be assumed from the model alone; it must be assessed through piloting, expert review, and empirical evaluation.

3. Proposed Methodological Framework

3.1. The GVEAR learning cycle

The proposed framework is called GVEAR: Generate–Verify–Evaluate–Argue–Reflect. It is a lesson-design sequence rather than a claim that all learning must occur in exactly five linear steps. A teacher may revisit a stage when students discover new information or need further language support.

Stage	Teacher/learner action	Critical-thinking focus	Evidence of learning
Generate	Learners ask GenAI for a short text, explanation, or set of arguments.	Clarifying the question and purpose.	Prompt specifies task, audience, and constraints.
Verify	Learners identify claims or language that require checking and consult teacher-approved sources.	Evidence checking and accuracy.	Learner records claim, source, and verification decision.
Evaluate	Learners compare AI output with evidence and identify omissions, bias, ambiguity, or errors.	Analysis and evaluation.	Annotated response with reasons for accepting/rejecting claims.
Argue	Learners state and support their own position orally or in writing.	Inference, explanation, and justification.	A reasoned paragraph or short presentation.
Reflect	Learners explain what they changed, what they learned, and how AI was used.	Self-regulation and responsible use.	Reflection note and transparent acknowledgement of AI assistance.

3.2. Example classroom task

Topic: Environmental choices and plastic waste (suggested for an age-appropriate intermediate-level class). Language aims: expressing opinions, using reasons and examples, and employing connectors such as because, however, and therefore. Critical-thinking aim: distinguishing a claim from evidence and identifying an unsupported generalisation.

Step 1—Generate. In a teacher-supervised setting, students request a short 120-word text giving three reasons to reduce single-use plastic. Students retain the prompt and output for discussion.

Step 2—Verify. The teacher provides two or three age-appropriate sources selected in advance. Students underline factual claims in the generated text and check whether each claim is supported by the sources. Students should not be required to create online accounts or share personal data to complete the task.

Step 3—Evaluate. In pairs, students label claims as supported, unsupported, unclear, or contradicted by the supplied sources. They must write one sentence explaining each decision. The teacher explicitly teaches that confident wording is not proof of accuracy.

Step 4—Argue. Each student writes a 120–150-word paragraph answering: ‘Which action should our school take first to reduce plastic waste, and why?’ The paragraph must include a claim, at least two reasons, and evidence from the supplied sources. Students may use AI for brainstorming only if the teacher’s rules allow it; the final reasoning must be their own.

Step 5—Reflect. Students complete a short log: ‘One AI claim I checked was…’; ‘I accepted/rejected it because…’; ‘One change I made to my own work was…’; ‘Next time I will…’. The teacher assesses the reasoning and evidence rather than rewarding the mere presence of AI-generated text.

3.3. Teacher mediation and safeguards

The teacher's role includes selecting reliable sources, checking AI-generated materials before classroom use, explaining task expectations, providing language scaffolds, and monitoring whether learners can complete the reasoning independently. The teacher should not upload identifiable student information to an AI service. School policies, parental permissions where required, age restrictions, data-protection rules, accessibility, and equitable access must be checked before implementation. UNESCO (2023) emphasises human agency, privacy, and age-appropriate, pedagogically justified use of GenAI.

4. Proposed Empirical Evaluation

This section is a research protocol for a future study, not a report of a completed experiment. No students have been represented as participants in the present manuscript, and no outcomes are claimed.

4.1. Design and setting

A feasible design would be a quasi-experimental pre-test/post-test study involving two existing Grade 8 English classes in a Tashkent secondary school, subject to school approval. One class would receive the GVEAR intervention and the other would continue with the school's usual English instruction for the same period. If class allocation cannot be randomised, baseline differences and the limitations of non-equivalent groups must be reported. The final paper must state the actual school, dates, number of participants, inclusion criteria, and any attrition.

A pilot could run for six weeks, with two 45-minute lessons per week incorporating one GVEAR task. The duration is a proposal and may be adjusted to the school timetable. Both groups should study the same curriculum themes and receive comparable instructional time; the planned difference should be the structured GenAI evaluation cycle, not extra teaching time.

4.2. Proposed instruments

Instrument A: a researcher-developed critical-thinking task set, including short claims, source excerpts, and a written justification task. Items should be reviewed by English-language teaching and assessment specialists before use. Instrument B: an AI-literacy scenario test covering output verification, prompt design, limitations, and responsible use. Instrument C: an English writing task scored with a rubric for task fulfilment, coherence, language control, and evidence-based reasoning. Instrument D: a learner reflection log and teacher observation checklist.

These instruments should not be described as validated until evidence of validity and reliability has been collected. At minimum, the researcher should obtain expert review, pilot the tasks with a comparable group where feasible, document revisions, and report inter-rater agreement for any rubric scored by more than one assessor.

4.3. Proposed rubric

Dimension	0 — Not demonstrated	1 — Emerging	2 — Demonstrated
Claim identification	Does not identify the relevant claim.	Identifies a claim with prompting or partial accuracy.	Accurately identifies the claim.
Evidence checking	Provides no checkable evidence.	Uses evidence but link to claim is unclear.	Uses relevant evidence and explains the link.

Dimension	0 — Not demonstrated	1 — Emerging	2 — Demonstrated
Evaluation	Accepts/rejects output without reason.	Gives a general reason.	Explains a specific strength, weakness, or uncertainty.
Argumentation	Opinion only; no supporting reason.	Gives a reason but weak support.	Gives a clear claim with relevant reasons/evidence.
Reflection/AI responsibility	Does not explain AI use or limitations.	Mentions a limitation or use decision vaguely.	Clearly explains verification, revision, and responsible use.

The rubric is a draft instrument for piloting, not a validated scale. If five dimensions are scored from 0 to 2, the total ranges from 0 to 10. The researcher should specify whether dimensions are equally weighted and whether the rubric is used as a formative classroom tool or a research outcome.

4.4. Procedure

Week 1: obtain approvals, inform families and students as required, pilot the instruments, and administer the baseline assessment. Weeks 2–7: deliver the planned lessons, keep a record of attendance and task completion, and collect anonymised student work. Week 8: administer parallel post-test tasks and collect reflections. If feasible, a delayed follow-up assessment may be added to examine retention. The schedule is illustrative and must be adapted to the actual study.

4.5. Data analysis plan

The analysis should begin with descriptive statistics (mean, standard deviation, score distributions, and missing data). The researcher should inspect baseline comparability and test assumptions before selecting inferential tests. Depending on the design and data, an analysis of covariance with post-test score as the outcome and baseline score as a covariate may be appropriate; a group-by-time model may also be considered. If assumptions are not met, robust or non-parametric alternatives should be justified. Report effect sizes and confidence intervals alongside p-values. For qualitative reflection logs, use a transparent coding scheme, provide examples that do not identify students, and describe how coding decisions were checked. The statistical method must be selected after the actual data and design are known, not chosen to produce a desired result.

4.6. Ethics and data management

Because participants are minors, the study requires approval through the relevant school and institutional procedures, informed parental/guardian consent where required, age-appropriate student assent, and clear information about voluntary participation. Participation or refusal must not affect grades. Data should be minimised, pseudonymised, stored securely, and reported in aggregate. The researcher should document the AI tool and version used, the prompts and output relevant to the intervention, and any safeguards for inaccurate or inappropriate output. Any applicable school, national, and platform rules must be followed.

5. Expected Contribution and Discussion

The proposed framework offers a practical way to connect English-language objectives with explicit evaluation behaviours. Its central premise is that the educational value of GenAI depends partly on the cognitive work required of learners. Generation alone may produce a text, but verification and evaluation require students to inspect claims; argumentation requires them to express a defensible position; and reflection makes the use of AI visible and open to critique.

The framework also offers teachers observable evidence for formative assessment. Instead of asking only whether a student completed a task, a teacher can inspect the prompt, the verification record, the comparison of claims, the student's argument, and the reflection log. This makes the learner's reasoning process more visible. The sequence may also support English practice because students read generated and source texts, discuss differences, and produce evidence-based writing.

Nevertheless, several limitations should be recognised. First, the framework is conceptual and has not yet been empirically tested in the present manuscript. Second, student performance may be affected by English proficiency, prior digital experience, teacher expertise, access to devices, and the quality of AI output. Third, a short intervention may capture task-specific performance rather than durable critical-thinking development. Fourth, a researcher-developed rubric may not have established psychometric properties. These issues should be addressed through piloting, transparent reporting, comparison groups where feasible, and replication across classes or schools.

The proposed novelty should therefore be stated cautiously: the article organises AI-supported English tasks into a five-stage cycle and aligns each stage with observable critical-thinking and AI-literacy indicators. Whether this constitutes a sufficiently original contribution for a target journal must be assessed through a systematic literature search and the journal's requirements. It should not be described as the first such model without evidence.

6. Conclusion

This article has proposed the GVEAR framework—Generate, Verify, Evaluate, Argue, and Reflect—for integrating generative AI into secondary-school English language teaching. The framework is intended to support critical thinking and AI literacy by requiring learners to verify claims, compare generated content with evidence, justify their own positions, and reflect on the responsible use of AI. It translates broad educational aims into a sequence of classroom actions and draft assessment indicators.

The article's contribution is methodological rather than empirical. A quasi-experimental evaluation has been outlined, including a possible intervention schedule, instruments, rubric, analysis plan, and ethical safeguards. The effectiveness of the framework cannot be concluded until the intervention is implemented and data are analysed transparently. Future work should pilot the tasks, obtain expert feedback on the instruments, report actual participant and outcome data, and examine how the framework performs across different proficiency levels and school contexts.

References

1. Facione, P. A. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report). California Academic Press. ERIC Document ED315423.
2. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274
3. Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. RELC Journal, 54(2), 537–550. https://doi.org/10.1177/00336882231162868
4. Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727
5. Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO. https://doi.org/10.54675/EWZM9535
6. Liu, J., Sihes, A. J. B., & Lu, Y. (2025). How do generative artificial intelligence (AI) tools and large language models (LLMs) influence language learners' critical thinking in EFL education? A systematic review. Smart Learning Environments, 12, Article 48. https://doi.org/10.1186/s40561-025-00406-0
7. Deng, L., & Jamaludin, K. A. (2026). Roles of generative artificial intelligence (GenAI) in English as a foreign language (EFL) instruction: A systematic literature review. SAGE Open. https://doi.org/10.1177/21582440261418315
8. Zhang, Y., Qiu, X., & Wang, J. (2026). A systematic research synthesis of generative AI literacy in English language education. Discover Artificial Intelligence, 6, Article 1074. https://doi.org/10.1007/s44163-026-02178-z
9. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71
10. Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1–16). ACM. https://doi.org/10.1145/3313831.3376727
11. Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press.
12. Bloom, B. S. (Ed.). (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain. David McKay.
13. Ennis, R. H. (2011). The nature of critical thinking: An outline of critical thinking dispositions and abilities. University of Illinois.
14. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—Where are the educators? International Journal of Educational Technology in Higher Education, 16, Article 39. https://doi.org/10.1186/s41239-019-0171-0