MAJESTY
MASPUL JOURNAL OF ENGLISH STUDIES
ISSN *2657-0157* (Online)

TYPE: Research Article
PUBLISHED: 12 *September* 2026
Vol. 8 No. 2, 2026; pp. 139-154
DOI: https://doi.org/10.33487/majesty.v8i2.490

ArguMap-AI: An Intelligent Claim–Evidence Mapping System for EFL Academic Writing

Mustakmi[1], Ismail[1], Sri Rosmiana[1], Ita Sarmita Samad[2]
[1]Universitas Muhammadiyah Enrekang, Indonesia
[2]Universitas Negeri Makassar, Indonesia

check for updates

Article History:
Received: June 2026
Revised: July 2026
Accepted: September 2026

Correspondence Author:
Ismail
ismail@unimen.ac.id
Universitas Muhammadiyah Enrekang

CITATION:
Mustakim, M., Ismail, I., Rosmiana, S., & Samad, I. S. (2026). ArguMap-AI: An Intelligent Claim–Evidence Mapping System for EFL Academic Writing. *Majesty Journal*, *8*(2), 139–154. https://doi.org/10.33487/majesty.v8i2.490

ABSTRACT

Background. Artificial intelligence is increasingly used to support EFL academic writing; however, most existing tools emphasize language correction, text generation, or general feedback rather than helping students examine the underlying structure of argumentative reasoning. In particular, EFL learners often experience difficulty connecting claims with relevant and sufficient evidence.

Purpose. This study aimed to develop and evaluate ArguMap-AI, an intelligent claim–evidence mapping system designed to support argumentative writing by identifying argumentative components, visualizing claim–evidence relationships, and providing diagnostic feedback for revision.

Methods. The study employed a design and development research approach supported by mixed-method evaluation. Participants were 21 fifth-semester students enrolled in the Writing Project course in the English Education Study Program at Universitas Muhammadiyah Enrekang during the 2025–2026 academic year. Data were collected through initial and revised argumentative drafts, an analytic writing rubric, ArguMap-AI-generated maps and feedback records, student perception questionnaires, and semi-structured interviews. Quantitative data were analyzed descriptively, while qualitative data were examined thematically.

Results. The findings showed improvement across all assessed dimensions of argumentative writing, with the largest gains occurring in claim–evidence alignment, reasoning and explanation, and evidence relevance and sufficiency. The mean total writing score increased from 18.19 to 26.33. Students generally perceived the visual mapping and diagnostic feedback as useful for identifying unsupported claims and improving revision decisions. Challenges included occasional component misclassification, complex visual maps in longer texts, and the continuing need for lecturer guidance.

Conclusion. ArguMap-AI shows potential as an intelligent reasoning scaffold for EFL academic writing by helping students visualize, evaluate, and revise argumentative relationships while preserving learner decision-making.

KEYWORDS
Academic writing; argument mining; artificial intelligence; claim–evidence mapping; EFL writing; intelligent writing support

INTRODUCTION

Academic writing is a central component of English as a Foreign Language (EFL) education in higher education because it requires learners to transform linguistic knowledge into coherent, evidence-supported academic arguments. Producing an effective academic text involves more than grammatical accuracy, appropriate vocabulary, or adherence to conventional essay structure. Writers must formulate defensible claims, select relevant evidence, establish logical relationships between claims and evidence, anticipate alternative perspectives, and revise arguments when existing support is insufficient. These demands are particularly challenging for

EFL learners because linguistic production and argumentative reasoning must often be managed simultaneously. Consequently, supporting EFL academic writing requires instructional approaches that address not only the surface accuracy of language but also the underlying architecture of students' reasoning.

Technology-assisted writing support has increasingly been introduced to address these challenges. Automated Writing Evaluation (AWE) systems can provide rapid and repeated feedback on student writing, allowing learners to revise without depending exclusively on lecturer feedback. Meta-analytic evidence indicates that AWE can positively influence writing quality, with particularly meaningful benefits reported for post-secondary EFL and ESL learners. Zhai and Ma (2023), in a meta-analysis of 26 studies involving 2,468 participants, reported a substantial overall effect of AWE on writing quality. Similarly, Ngo et al. (2024) found positive effects of AWE on EFL/ESL writing performance and showed that its effectiveness can vary according to learner characteristics, duration, feedback focus, and instructional implementation. These findings establish automated feedback as a potentially valuable component of contemporary writing pedagogy.

The rapid development of generative artificial intelligence (GenAI) has considerably expanded these possibilities. Unlike traditional AWE systems, contemporary AI tools can generate explanations, suggest revisions, interact dialogically with learners, and provide feedback on multiple aspects of writing. Recent studies have reported benefits of AI-generated feedback for language accuracy, revision practices, learner independence, and overall writing performance (Han & Li, 2024; Mekheimer, 2025; Alnemrat et al., 2025; Mujeeb et al., 2026). Indonesian students have also reported valuing AI feedback for its immediacy and usefulness, particularly for language-related aspects of EFL writing (Dewi et al., 2026). These developments indicate that AI is increasingly becoming part of the ecology of university writing instruction rather than merely an optional editing tool.

Nevertheless, the availability of increasingly sophisticated AI feedback does not necessarily resolve one of the most fundamental problems in argumentative writing: whether a claim is adequately supported by appropriate evidence and reasoning. A grammatically polished text can still contain unsupported assertions, evidence that does not logically support the thesis, weak relationships between ideas, or arguments that ignore relevant counterpositions. Research on students' engagement with ChatGPT feedback provides evidence for this distinction. Cengiz et al. (2025) found that EFL learners were more likely to accept language-level feedback than feedback concerning content and organization because higher-level revisions demanded greater interpretation and critical judgment. Han and Li (2024) similarly emphasized the value of integrating AI analysis with teacher mediation rather than treating automated feedback as self-sufficient. Thus, producing more feedback does not automatically result in deeper argumentative development.

This issue becomes increasingly important as GenAI systems generate fluent and seemingly sophisticated academic prose. The educational challenge is therefore shifting from whether AI can improve textual form toward whether AI can help students see, evaluate, and improve their own reasoning. Recent research on AI-assisted feedback literacy supports this shift. Liu et al. (2026) found that GenAI feedback can strengthen students' ability to interpret, evaluate, and act on feedback, but they also identified continuing challenges involving cognitive load, disciplinary knowledge, and academic conventions. In other words, the educational value of AI depends on whether learners remain cognitively involved in evaluating and revising their writing rather than simply accepting machine-generated language.

One promising approach to making reasoning more visible is argument mapping. Argument mapping represents argumentative relationships visually, typically distinguishing a central claim or thesis from supporting reasons, evidence, counterclaims, and rebuttals. By externalizing these relationships, a writer can inspect an argument as a connected structure rather than as a sequence of sentences. A systematic review by Nesbit and Liu (2025), covering 124 studies in higher education, demonstrated substantial scholarly interest in argument mapping, including its relationships with critical thinking, collaborative learning, and the quality and structure of arguments. The review indicates that argument visualization has considerable potential as a learning design strategy in postsecondary education.

Within EFL writing, computer-aided argument mapping has also produced encouraging results. Robillos and Thongpai (2022) reported that integrating computer-assisted argument mapping with a metacognitive approach supported EFL students' argumentative writing performance, including content and coherence, as well as self-regulated learning. Alsmari (2022) similarly found gains in argumentative writing and self-regulation among university EFL learners using computer-aided argument mapping. These studies suggest that making argumentative relationships visually explicit can help learners organize ideas before and during writing rather than relying solely on linear textual drafting.

More recently, argument mapping has begun to intersect with AI-supported learning. Chen et al. (2025) examined online argumentation supported by ChatGPT and argument maps and highlighted the importance of clear argumentative structures and timely guidance for developing students' critical thinking. In the specific context of EFL writing, Gu et al. (2026) reported that collaborative argument mapping supported global cohesion and helped learners perceive relationships among claims, reasons, and evidence more clearly. Students also reported gains in critical thinking and more deliberate writing processes, although information load and unequal attention to linguistic form remained challenges. These findings point toward the value of combining visual argument structures with intelligent support while simultaneously showing that conventional argument mapping still places substantial responsibility on learners to identify and classify argumentative components manually.

A complementary technological development is argument mining, a field of natural language processing concerned with computationally identifying argumentative components and the relationships among them. Educational argument-mining systems can potentially detect claims, premises, supporting relations, counterarguments, and other structures in students' texts, thereby creating opportunities for automated diagnosis and feedback. Pereira et al. (2025), in a comprehensive survey of argument mining in education, concluded that these technologies have substantial potential for assessing assignments, understanding students' positions, and supporting teaching and learning. However, the authors also identified continuing gaps involving educational applications, linguistic diversity, representation of argument structures, and the translation of computational analysis into meaningful pedagogical support.

A further difficulty is that many argument-mining technologies have been designed or evaluated using completed texts, whereas students require support while their arguments are still developing. Schaller et al. (2025) demonstrated that argument-mining models that perform well on complete essays can become less reliable when applied to incomplete learner texts. Their study showed that limited or changing context influences model performance and argued for systems capable of operating within dynamic writing processes. This limitation is particularly relevant to educational writing support because identifying a weak argument only after an essay has been completed provides fewer opportunities for formative intervention than identifying weaknesses during drafting and revision.

The literature therefore reveals an important gap at the intersection of AI-assisted EFL writing, argument mapping, and educational argument mining. AI writing research has generated extensive evidence concerning automated feedback, revision, language accuracy, and learner perceptions. Argument-mapping research has demonstrated the value of visualizing argumentative structures. Argument-mining research has developed computational techniques for identifying argumentative elements. However, comparatively limited attention has been directed toward an EFL writing system that integrates these capabilities to automatically identify students' own argumentative components, visualize the relationship between claims and supporting evidence, and provide formative guidance that encourages students to revise the reasoning rather than having AI rewrite the text for them.

This gap is also relevant to the Indonesian university context. Recent evidence indicates growing acceptance and use of AI feedback among Indonesian EFL students, while also showing that AI and human feedback may serve different pedagogical purposes (Dewi et al., 2026). Nevertheless, much of the rapidly expanding AI-writing literature continues to focus on students' perceptions, language feedback, general writing performance, and the use of existing generative platforms. The development and classroom evaluation of a specialized intelligent system focused specifically on claim–evidence relationships and argumentative reasoning remain relatively uncommon, especially within English teacher education.

The present study addresses this problem through the development of ArguMap-AI: An Intelligent Claim–Evidence Mapping System for EFL Academic Writing. ArguMap-AI is conceived not primarily as a grammar checker, essay generator, or automatic rewriting system, but as an intelligent writing-support environment designed to make students' argumentative reasoning visible. At its core, the system analyzes argumentative drafts, identifies relevant argumentative components, and maps relationships between claims and evidence so that learners can recognize unsupported claims, insufficient evidence, and problematic claim–evidence alignment. The intended function of the technology is therefore diagnostic and formative: AI assists students in examining their own reasoning while the responsibility for evaluating, developing, and revising the argument remains with the student.

The conceptual novelty of ArguMap-AI lies in this shift from AI as text producer to AI as reasoning scaffold. Existing generative systems can readily produce fluent sentences or complete argumentative paragraphs, potentially obscuring whether students themselves understand the underlying relationships among propositions. By contrast, ArguMap-AI seeks to expose those relationships and return them to learners for evaluation. Such positioning is particularly relevant in light of recent EFL argument-mapping research showing that visualized relationships among claims, reasons, and evidence can encourage more deliberate writing processes (Gu et al., 2026), and research demonstrating that students engage differently with feedback depending on the cognitive demands involved (Cengiz et al., 2025).

The study is situated in the English Education Study Program at Universitas Muhammadiyah Enrekang, South Sulawesi, Indonesia, and involves fifth-semester undergraduate students in the 2025–2026 academic year. This setting provides a contextual basis for examining how an intelligent claim–evidence mapping system operates with university EFL writers and how learners respond when AI is used to analyze rather than replace their argumentative work. Locating the study within an English Education program is also important because the participants are developing both academic English competence and experience with pedagogical practices that may subsequently inform their professional engagement with language education.

Accordingly, this study aims to develop and evaluate ArguMap-AI as an intelligent claim–evidence mapping system for EFL academic writing. The study focuses on the system's ability to identify and visualize argumentative relationships in students' writing, its contribution to the development and revision of argumentative texts, students' experiences of using the system, and the technological and pedagogical challenges encountered during implementation. By bringing together AI-assisted writing, argument visualization, and educational argument mining, the study seeks to contribute a more reasoning-oriented perspective to intelligent EFL writing support—one in which technology assists learners not simply to produce better-looking texts, but to construct more visible, evidence-based, and defensible academic arguments.

The study addresses the following research questions:

RQ1: How can ArguMap-AI be designed to identify and visually map claim–evidence relationships in the argumentative writing of fifth-semester EFL students?

RQ2: To what extent does the use of ArguMap-AI support students in improving claim development, evidence use, claim–evidence alignment, and overall argumentative writing quality?

RQ3: How do fifth-semester students of the English Education Study Program perceive the usefulness, usability, and feedback value of ArguMap-AI during argumentative writing and revision?

RQ4: What technological and pedagogical challenges emerge during the implementation of ArguMap-AI in the English Education Study Program at Universitas Muhammadiyah Enrekang?

METHOD

Design

This study employed a design and development research approach supported by mixed-method evaluation to develop and examine ArguMap-AI, an intelligent claim–evidence mapping system for EFL academic writing. The developmental component focused on translating pedagogical requirements of argumentative writing into a functional prototype, while the evaluative component investigated the system's performance, its contribution to students' argumentative writing, students' perceptions of its usefulness and usability, and challenges encountered during classroom implementation.

The study was conducted through five interrelated stages: needs analysis, system design, prototype development, classroom implementation, and evaluation. This structure enabled the technological development of ArguMap-AI to remain connected to authentic instructional needs rather than being treated as an independent software-development activity.

Quantitative data were used to examine changes in selected dimensions of students' argumentative writing, whereas qualitative data were used to explain how students interacted with the system and how they perceived its feedback. This combination was considered appropriate because the study addressed both the technical-pedagogical development of the system and its use within an actual EFL writing environment.

Participants and Setting

The study was conducted in the English Education Study Program at Universitas Muhammadiyah Enrekang, South Sulawesi, Indonesia, during the 2025–2026 academic year. The participants were 21 fifth-semester undergraduate students enrolled in the Writing Project course. The participants constituted an intact classroom group and were selected because the course required students to engage in advanced writing activities involving idea development, claim formulation, evidence selection, argumentative organization, revision, and the production of extended academic texts.

The Writing Project course provided an appropriate setting for evaluating ArguMap-AI because students were expected not only to produce linguistically acceptable academic writing but also to construct coherent and evidence-supported arguments. This context allowed the study to examine how an intelligent claim–evidence mapping system could assist students in identifying argumentative components, evaluating relationships between claims and evidence, and revising weaknesses in their argumentative structure.

All 21 students participated in the instructional activities involving ArguMap-AI. Participation in the research component was voluntary, and students were informed that their decision to participate or not participate would not affect their academic grades. To maintain confidentiality, participant identities were anonymized using codes such as S1–S21 throughout data collection, analysis, and reporting.

ArguMap-AI System Development

The development of ArguMap-AI began with an analysis of common difficulties encountered in EFL argumentative writing, particularly unsupported claims, insufficient evidence, weak claim–evidence alignment, and limited awareness of argumentative structure. These needs informed the functional design of the system.

At the prototype level, ArguMap-AI consisted of four main functional components: (1) Text Input Module, through which students submitted argumentative drafts; (2) Argument Identification Module, which detected relevant argumentative components in the submitted text; (3) Claim–Evidence Mapping Module, which visually represented relationships between identified claims and supporting evidence; and (4) Formative Feedback Module, which highlighted potential structural weaknesses and prompted students to reconsider unsupported or insufficiently connected arguments.

The system was deliberately designed not to generate complete replacement essays. Instead, its principal function was to make students' argumentative reasoning visible and provide diagnostic prompts that could support revision. The technical architecture and computational mechanisms were documented separately from the pedagogical reporting of the study.

Instruments

Five sources of data were used: students' argumentative drafts, an argumentative-writing assessment rubric, system-generated claim–evidence maps, a student perception questionnaire, and semi-structured interviews.

Argumentative Writing Tasks

Students completed an argumentative writing task before and after the classroom implementation. The tasks required students to formulate a clear position, provide appropriate supporting evidence, explain relationships between evidence and claims, and organize arguments coherently. Initial and revised drafts were retained to examine changes in argumentative structure and writing quality.

Argumentative Writing Rubric

Students' texts were assessed using an analytic rubric focusing on seven dimensions: (1) claim clarity, referring to the extent to which the writer presents a clear, specific, and defensible position; (2) relevance and sufficiency of evidence, examining whether the evidence used is appropriate, credible, and adequate to support the argument; (3) claim–evidence alignment, assessing the logical relationship between each claim and the evidence provided; (4) reasoning and explanation, evaluating how effectively the writer explains why the evidence supports the claim; (5) counterargument and rebuttal, assessing the writer's ability to acknowledge alternative perspectives and respond to them appropriately; (6) organization and coherence, examining the logical arrangement and connection of ideas across the text; and (7) overall argumentative quality, representing the overall strength, persuasiveness, and consistency of the argument. The rubric enabled systematic comparison between students' writing before and after using ArguMap-AI.

System-Generated Records

ArguMap-AI automatically produced visual maps and feedback records from students' drafts. These records documented identified argumentative components and served as evidence of how students interacted with system-generated analysis during revision.

Student Perception Questionnaire

A questionnaire was administered after implementation to examine students' perceptions of usefulness, usability, feedback clarity, and perceived contribution to writing revision. Closed-ended items provided descriptive data, while open-ended questions allowed students to explain benefits, difficulties, and suggested improvements.

Semi-Structured Interviews

Semi-structured interviews were conducted with [number] purposively selected students representing different patterns of system use and classroom participation. Questions explored how students interpreted the claim–evidence maps, responded to feedback, revised their drafts, and experienced technological or pedagogical difficulties.

Data Collection Procedure

Data collection was organized sequentially to ensure that students' initial writing performance, interaction with ArguMap-AI, and post-implementation responses could be documented systematically. The procedure was designed to capture both the development of students' argumentative writing and their experiences in using the system, while maintaining consistency between the instructional activities, research instruments, and research questions.

Phase 1: Baseline Writing and System Orientation

Students first completed an argumentative writing task without ArguMap-AI support. The initial drafts provided baseline evidence concerning claim development, evidence use, and argumentative organization. Students were then introduced to the purpose and functions of ArguMap-AI and received guided practice in interpreting claim–evidence maps and feedback.

Phase 2: ArguMap-AI-Assisted Writing and Revision

Students used ArguMap-AI during subsequent argumentative-writing activities. The general workflow was:

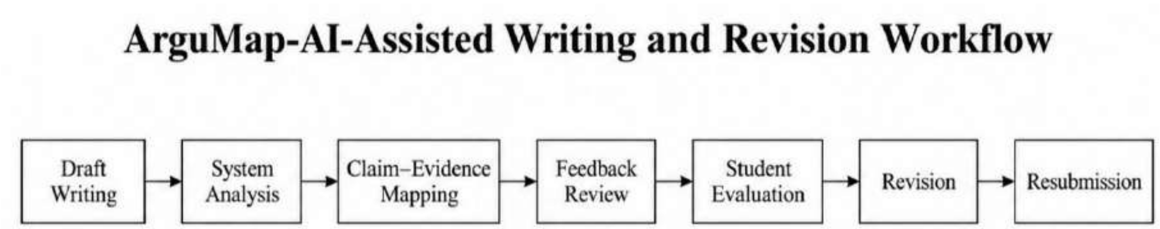

Figure 1. Workflow of ArguMap-AI-assisted writing and revision

Students were encouraged to examine the system feedback critically rather than automatically accept every recommendation. Lecturer support was provided when students experienced difficulty interpreting maps or determining how feedback should influence revision. Initial drafts, system-generated maps, feedback records, and revised drafts were retained for analysis.

Phase 3: Evaluation

At the end of implementation, students completed a final argumentative writing task and the perception questionnaire. Selected students subsequently participated in semi-structured interviews. The interviews were used to explore experiences that could not be fully captured by writing scores or questionnaire responses.

Data Analysis

Quantitative and qualitative data were analyzed separately and subsequently integrated. Students' initial and final argumentative texts were scored using the analytic rubric. Descriptive statistics, including mean scores, standard deviations, and score differences, were calculated for each assessed dimension and for overall argumentative-writing quality. Where the final dataset met appropriate assumptions, a paired statistical comparison could be conducted to examine differences between pre- and post-implementation scores. System-generated records were examined to identify recurring patterns such as unsupported claims, insufficient evidence, weak claim–evidence alignment, and revisions made after feedback.

Questionnaire responses were summarized using frequencies, percentages, and mean ratings. Open-ended questionnaire responses and interview transcripts were analyzed thematically. Initial codes were grouped into broader themes related to perceived usefulness, feedback interpretation, revision practices, usability, learner autonomy, technological limitations, and pedagogical challenges. Data from writing artifacts, system records, questionnaires, and interviews were triangulated to determine whether improvements visible in students' texts were consistent with their experiences of using ArguMap-AI.

RESULTS

ArguMap-AI Claim–Evidence Mapping and Feedback

ArguMap-AI was implemented as an intelligent writing-support environment that processed students' argumentative drafts and produced visual representations of relationships between claims and supporting evidence. The system workflow enabled students to submit a draft, inspect the generated map, review diagnostic feedback, evaluate whether the feedback was appropriate, revise their writing, and resubmit the text.

During implementation, the system most frequently identified four argumentative problems: unsupported claims, insufficient or irrelevant evidence, weak claim–evidence alignment, and incomplete reasoning between claims and evidence. It also identified instances in which counterarguments or rebuttals were absent from otherwise developed argumentative texts.

Across the 21 initial drafts, 74 claim units and 61 evidence units were identified. Twenty-six claims were initially presented without adequate supporting evidence, while 29 claim–evidence relationships were marked as weak or insufficiently explicit. Thirty-one instances showed evidence that was present in the text but inadequately explained in relation to the corresponding claim.

Following system-supported revision, the number of identified evidence units increased, whereas unsupported and weakly connected claims decreased.

Table 1 *Changes in Argumentative Components Identified by ArguMap-AI*

Argumentative Component	Initial Drafts	Revised Drafts
Claims identified	74	79
Evidence units identified	61	83
Unsupported claims	26	8
Weak claim–evidence relationships	29	10
Insufficient reasoning/explanation	31	12
Drafts without counterargument/rebuttal	15	6

The system-generated records also documented **166 formative feedback prompts** during the revision process. The largest proportion addressed claim–evidence alignment, followed by evidence sufficiency and reasoning.

Table 2 *Types of ArguMap-AI Feedback Generated During Revision*

Feedback Focus	Frequency	Percentage
Claim–evidence alignment	52	31.3%
Evidence relevance/sufficiency	41	24.7%
Reasoning and explanation	38	22.9%
Counterargument/rebuttal	21	12.7%
Organization/coherence	14	8.4%
Total	**166**	**100%**

Students acted fully on 133 feedback prompts (80.1%), partially responded to 21 prompts (12.7%), and did not adopt 12 suggestions (7.2%). The rejected suggestions were retained as part of the data because students were instructed to evaluate rather than automatically accept system feedback.

A typical system interaction occurred when a student presented a clear claim but attached evidence that did not directly justify the claim. ArguMap-AI visually marked the connection as weak and generated a diagnostic prompt asking the student to reconsider how the evidence supported the stated position. After reviewing the map, the student either strengthened the explanation, replaced the evidence, or modified the claim.

Changes in Students' Argumentative Writing

The second research question examined the extent to which students' argumentative writing changed after using ArguMap-AI. Each writing sample was assessed across seven dimensions using a five-point analytic scale: claim clarity, evidence relevance and sufficiency, claim–evidence alignment, reasoning and explanation, counterargument and rebuttal, organization and coherence, and overall argumentative quality. Descriptive results showed higher mean scores on all seven dimensions after implementation (table 3).

Table 3 *Students' Argumentative Writing Scores Before and After ArguMap-AI Use (N = 21)*

Dimension	Initial M	Initial SD	Revised M	Revised SD	Mean Gain
Claim clarity	2.95	0.59	3.86	0.79	0.90
Evidence relevance and sufficiency	2.57	0.51	3.81	0.75	1.24
Claim–evidence alignment	2.38	0.50	3.76	0.62	1.38
Reasoning and explanation	2.48	0.51	3.81	0.87	1.33
Counterargument and rebuttal	2.24	0.54	3.43	0.60	1.19
Organization and coherence	3.05	0.59	4.00	0.77	0.95
Overall argumentative quality	2.52	0.75	3.67	0.91	1.14

The mean total score increased from 18.19 (SD = 1.25) in the initial drafts to 26.33 (SD = 1.93) in the revised drafts, representing an average increase of 8.14 points on the 35-point rubric.

The greatest improvement was observed in claim–evidence alignment (M gain = 1.38), followed by reasoning and explanation (M gain = 1.33) and evidence relevance and sufficiency (M gain = 1.24). Counterargument and rebuttal remained the lowest-scoring dimension after revision despite showing improvement.

The draft comparisons also revealed qualitative changes in how students structured arguments. Initial texts frequently contained claims followed immediately by examples or factual statements without explicitly explaining why the evidence supported the position. Revised drafts more frequently included a sequence of claim → evidence → explanation, while some students additionally incorporated counterarguments and rebuttals.

For example, an initial claim from one writing sample stated:

> "Social media has a negative effect on university students because it disturbs their study."

After the system identified insufficient evidence and a weak claim–evidence connection, the revised passage included a more specific claim, relevant evidence, and an explicit explanation of how the evidence supported the position. The revision represented a change not only in linguistic form but also in the structural relationship among argumentative components.

A similar pattern appeared when students used evidence. In several initial drafts, sources were inserted after claims with limited explanation. Following ArguMap-AI feedback, students more frequently added interpretive statements linking the cited information to the argumentative position.

Students' Perceptions of ArguMap-AI

Students' perceptions were examined in relation to usefulness, usability, feedback clarity, support for revision, and learner control. Overall questionnaire responses were positive (table 4).

Table 4 *Students' Perceptions of ArguMap-AI (N = 21)*

Dimension	Mean	SD	Agree/Strongly Agree
Usefulness for understanding argument structure	4.48	0.51	90.5%
Helpfulness for identifying unsupported claims	4.43	0.60	85.7%
Helpfulness for evaluating claim–evidence relationships	4.52	0.51	90.5%
Feedback clarity	4.24	0.62	85.7%
Ease of use	4.19	0.68	81.0%
Support for writing revision	4.43	0.51	90.5%
Preservation of student decision-making	4.29	0.56	90.5%
Intention to use the system again	4.33	0.58	85.7%

Note. Responses were recorded on a five-point scale from 1 = strongly disagree to 5 = strongly agree.

The highest mean score was obtained for the system's ability to help students evaluate claim–evidence relationships (M = 4.52), followed by understanding argument structure (M = 4.48).

Ease of use received the lowest mean score (M = 4.19), although the rating remained positive.

Open-ended responses showed that 18 of the 21 students considered the visual map useful for identifying the structure of their arguments. Seventeen reported that the system helped them notice unsupported claims, and 16 indicated that they had previously focused more strongly on grammar than on relationships between ideas.

One student explained:

> "Before using the map, I usually checked grammar and vocabulary first. The system showed me that some sentences looked good, but the evidence did not really support my claim." (S6)

Another participant described the visual component as particularly useful:

> "When I saw the claim and evidence separately on the map, I could see which part was missing. In the essay itself, I did not notice it clearly." (S12)

Students also reported that system feedback did not always need to be followed automatically:

> "Sometimes I did not agree with the feedback, so I checked my paragraph again and decided whether I needed to change it. I think that was useful because the AI did not write the answer for me." (S16)

The interview data similarly indicated that students generally perceived ArguMap-AI as a diagnostic tool rather than an automatic writing generator.

Revision Decisions Following ArguMap-AI Feedback

Analysis of initial and revised drafts revealed three main revision patterns. The first involved adding evidence to unsupported claims. Students either introduced new supporting information or removed claims that they could not sufficiently justify.

The second involved strengthening the explanation between existing evidence and claims. In these cases, students did not necessarily add new sources but clarified why the information provided was relevant to their position.

The third involved restructuring an argument. Several students revised thesis statements, reorganized supporting paragraphs, or introduced counterarguments after inspecting the visual representation of their essays.

Table 5 *Major Revision Actions Following ArguMap-AI Feedback*

Revision Action	Students Showing the Pattern (n)
Added or replaced supporting evidence	18
Strengthened claim–evidence explanation	17
Revised or narrowed a claim	14
Reorganized argumentative paragraphs	13
Added a counterargument	12
Developed or strengthened rebuttal	10
Rejected at least one AI suggestion after evaluation	9

The finding that nine students rejected at least one recommendation was noteworthy within the system records. These students generally retained their original wording or argumentative structure when they considered the AI interpretation inappropriate or when they could justify the existing relationship more clearly than the map initially represented.

Technological and Pedagogical Challenges

Although responses were generally positive, several technological and pedagogical challenges emerged from system records, open-ended questionnaires, classroom observations, and interviews.

The most frequently reported technological issue concerned misclassification of argumentative components. In some drafts, the system interpreted explanatory sentences as evidence or treated contextual statements as claims. Four students explicitly referred to this problem in questionnaire responses.

A second issue concerned the complexity of maps generated from longer paragraphs. Five students reported that visual displays became more difficult to interpret when many claims and evidence units were presented simultaneously.

One participant stated:

> "For a short paragraph, the map was very clear. When the essay became longer, there were many connections and I needed more time to understand them." (S9)

Feedback specificity represented another limitation. Six students indicated that some prompts correctly showed where a weakness existed but did not always explain sufficiently how the problem should be resolved.

Pedagogically, students also required initial guidance to distinguish evidence from reasoning.

Classroom records showed that some students initially expected any citation or example to function automatically as evidence. Lecturer clarification was therefore necessary to explain that evidence needed both relevance and an explicit connection to the claim.

Table 6 *Main Challenges Identified During ArguMap-AI Implementation*

Challenge	Evidence Across Data Sources
Occasional misclassification of claims/evidence	System records; 4 questionnaire responses
Complex visual maps in longer texts	5 questionnaire responses; interviews
Feedback occasionally too general	6 questionnaire responses
Difficulty distinguishing evidence from reasoning	Classroom observations; interviews
Initial tendency to accept AI feedback automatically	Early classroom observations
Additional time required for reviewing maps and revision	Questionnaire and interview responses
Need for lecturer guidance during early implementation	Classroom observations

A further classroom pattern involved students' initial assumption that automated feedback should be accepted because it originated from an AI system. During system orientation and subsequent activities, students were repeatedly encouraged to treat the feedback as a recommendation requiring evaluation. Later revision records showed greater variation in students' responses to system suggestions, including acceptance, modification, and rejection.

Overall, the results showed that the most visible changes after ArguMap-AI implementation occurred in aspects directly related to the system's central function: claim–evidence alignment, evidence use, and explanation of reasoning. The system records also showed that students did not respond uniformly to automated feedback; some suggestions were accepted, others were modified, and a smaller proportion were rejected after student evaluation. The questionnaire and interview data further indicated that students valued the ability to visualize the structure of their arguments while continuing to make their own revision decisions.

DISCUSSION

The present study developed and evaluated ArguMap-AI, an intelligent claim–evidence mapping system designed to support argumentative writing among fifth-semester students in the English Education Study Program at Universitas Muhammadiyah Enrekang. Four findings are particularly important. First, ArguMap-AI was able to represent students' argumentative structures by identifying claims, evidence, and potential weaknesses in the relationships between them. Second, students' revised writing showed improvement across all assessed dimensions, with the largest gains occurring in claim–evidence alignment, reasoning and explanation, and evidence relevance and sufficiency. Third, students generally perceived the system positively, particularly its capacity to make argumentative relationships visible and identify unsupported claims. Fourth, the implementation also revealed limitations involving occasional argument-component misclassification, visual complexity in longer texts, uneven feedback specificity, and the continuing need for lecturer mediation.

Making Argumentative Reasoning Visible

The most distinctive finding concerns the role of ArguMap-AI in making the internal structure of students' arguments visible. The system did not merely identify linguistic errors but separated argumentative components and displayed relationships between claims and evidence. This functionality appeared particularly relevant because initial drafts frequently contained claims that were unsupported or accompanied by evidence whose relationship to the claim remained implicit.

The finding supports previous research showing that argument visualization can strengthen students' awareness of argumentative structure. Nesbit and Liu (2025) reported that argument mapping in higher education has been associated with critical thinking and improvement in argument structure, while Robillos and Thongpai (2022) demonstrated that computer-aided argument mapping can support EFL argumentative writing and self-regulated learning. Similarly, Guo et al. (2022) introduced *Argumate*, a chatbot designed to scaffold argument construction, including idea generation and counterargument development. These studies collectively indicate that digital technologies can support argumentative writing when they make reasoning processes more accessible to learners.

ArguMap-AI, however, occupies a somewhat different instructional space. Rather than primarily functioning as an interactive conversational partner, its central function is to diagnose and visualize the relationship between components already produced by the student. This distinction is pedagogically significant because the system begins with students' own writing rather than generating argumentative content on their behalf. In this way, the technology functions less as an answer-generating mechanism and more as a mirror through which learners can inspect the structure of their reasoning.

This orientation also responds to concerns surrounding the increasing use of generative AI in second-language writing. Barrot (2023) notes that ChatGPT provides substantial opportunities for L2 writing support but also raises concerns related to accuracy, authorship, and overdependence. ArguMap-AI addresses this tension by maintaining students as the primary producers and decision makers: AI identifies potential structural problems, but students determine whether and how those problems should be revised.

Improvement in Claim–Evidence Alignment and Argument Quality

The strongest improvements occurred in claim–evidence alignment, followed by reasoning and explanation and evidence relevance and sufficiency. This pattern is noteworthy because these dimensions correspond directly to the central diagnostic functions of ArguMap-AI. Students were not simply exposed to generic writing advice; they repeatedly encountered visual representations indicating where a claim lacked evidence, where evidence was insufficiently connected to a claim, or where an explanatory bridge between the two was missing.

The finding is consistent with educational argument-mining research suggesting that automatically identifying argumentative units can support learners when computational analysis is transformed into actionable feedback (Pereira et al., 2025). It also corresponds with the development of systems such as FEAT-writing, which uses NLP-based argument analysis to generate targeted exercises and feedback for argumentative writing. Ding et al. (2025) reported general improvements in students' writing after interaction with different forms of automated argumentative feedback.

The present findings add an EFL-specific dimension to this line of work. For EFL students, weaknesses in argumentative writing may not arise solely because learners lack ideas. Students may possess relevant knowledge but fail to make the logical relationship between a claim and supporting evidence sufficiently explicit in English. A system capable of visualizing this relationship can therefore provide a form of external representation that reduces the need for learners to infer argumentative structure from a linear text alone.

The results also align with recent Indonesian evidence concerning human–AI collaborative writing. Pratama et al. (2025) reported substantial improvement in Indonesian EFL students' argumentative writing and critical-thinking dimensions following a structured human–AI collaborative writing strategy. Importantly, their findings support an approach in which artificial intelligence complements rather than replaces human reasoning.

However, the findings should not be interpreted as evidence that visualization itself automatically produces better reasoning. Students improved most clearly when they acted on the map, reconsidered evidence, explained argumentative relationships, and revised their texts. Thus, the pedagogically relevant mechanism appears to involve an interaction between automated diagnosis and learner revision rather than automated diagnosis alone.

From Automated Feedback to Learner Decision-Making

Another important finding was that students did not uniformly accept all ArguMap-AI recommendations. Although most feedback prompted revisions, nine students rejected at least one suggestion after reviewing it. This behavior is pedagogically meaningful because critical rejection of AI feedback may represent a stronger learning outcome than unquestioning acceptance.

Research on second-language writers' engagement with ChatGPT feedback demonstrates that learner uptake varies substantially according to feedback type. Cengiz et al. (2025) found that learners more readily accepted language-level feedback, whereas feedback concerning organization and content produced more varied responses, including rejection or no revision. Similarly, research on writing feedback literacy emphasizes that receiving feedback and effectively using feedback are distinct capacities (Weng et al., 2024). Students need to interpret, evaluate, and make judgments about information before deciding how it should affect their texts.

This distinction helps clarify the intended pedagogical role of ArguMap-AI. The system should not be judged solely by how frequently students accept its recommendations. A system that achieves near-total acceptance because students assume that AI is always correct could inadvertently reduce learner agency. By contrast, an intelligent writing environment should encourage students to ask whether the identified claim is genuinely a claim, whether the system has correctly classified a piece of evidence, and whether the proposed weakness requires textual revision.

This orientation also resonates with research on chatbot-assisted argumentative writing. Guo et al. (2024) showed that EFL learners' engagement with chatbot-assisted argumentation needs to be understood as a human–machine collaborative activity rather than a simple technology-to-user transmission process. Similarly, Guo and Li (2024) reported that personalized AI writing assistance can support EFL learners, but effective use depends on how students appropriate technological affordances for their own writing purposes. Accordingly, one contribution of ArguMap-AI may lie in promoting evaluative agency: students are required not simply to receive feedback but to judge its applicability to their intended argument.

Visual Feedback as a Cognitive Scaffold

Students' positive responses to visual claim–evidence maps provide another important finding. The highest perception score concerned the system's ability to help students evaluate claim–evidence relationships. Interview responses similarly indicated that argumentative weaknesses became easier to notice when claims and evidence were separated visually from the continuous prose of an essay.

This finding is supported by recent research on visualized GenAI feedback. Zou et al. (2025) found that visually enhanced AI feedback contributed to stronger coherence and cohesion in EFL writing while reducing negative emotional responses and cognitive load compared with text-only feedback. Their study emphasizes that the educational effectiveness of AI-generated feedback depends not only on what information is provided but also on how that information is represented.

The visual dimension of ArguMap-AI may be particularly useful for argumentative writing because the relationships being represented are inherently relational. Claim–evidence alignment is difficult to convey through isolated corrective comments such as *"provide more evidence"*. A map can instead enable students to observe that several claims converge on one piece of evidence, that one claim remains isolated, or that evidence exists but lacks a clear argumentative connection.

Nevertheless, the results also showed that the visual advantage decreased as texts became longer and maps became more complex. This suggests that visualization can become a source of cognitive overload when too many argumentative components and relationships appear simultaneously. Woo et al. (2024) similarly reported substantial cognitive demands when EFL learners used ChatGPT for writing, emphasizing the importance of carefully scaffolding technology-supported tasks. Future versions of ArguMap-AI should therefore consider progressive visualization, allowing students to inspect argument structures at different levels—for example, thesis-level relationships first, paragraph-level relationships second, and sentence-level details only when required.

ArguMap-AI as a Reasoning Scaffold Rather Than a Writing Generator

The findings provide support for the conceptual positioning of ArguMap-AI as a reasoning scaffold. This positioning is important because much contemporary AI-assisted writing research focuses on systems capable of generating or rewriting language. Although these affordances can support students' writing, they can also blur the boundary between learner-produced reasoning and machine-produced text.

Recent human–AI collaborative writing research demonstrates that the most educationally productive applications of AI may be those that preserve meaningful human intellectual involvement. Pratama et al. (2025) found that structured human–AI collaboration improved argumentative writing and critical thinking among Indonesian EFL learners. Guo et al. (2024) likewise found that AI-supported peer feedback enhanced feedback quality and students' writing ability, indicating that AI can function productively when embedded within activities requiring learners to make substantive judgments.

ArguMap-AI extends this principle by assigning different responsibilities to the human writer and the intelligent system. The student remains responsible for generating the argument, selecting evidence, evaluating diagnostic feedback, and deciding on revisions, while the system supports detection, relational analysis, visualization, and diagnostic prompting. This distribution of responsibility distinguishes ArguMap-AI from systems that generate complete replacement passages. The system's pedagogical objective is not to reduce the intellectual work of writing but to make that work more observable and manageable.

The Continuing Importance of Lecturer Mediation

Despite positive student perceptions and improvements in writing, the findings also demonstrate that AI did not eliminate the need for lecturer involvement. Students initially experienced difficulty distinguishing evidence from reasoning, interpreting complex maps, and deciding whether particular feedback should be accepted. Lecturer guidance was therefore particularly important during early implementation.

This finding is consistent with Han and Li's (2024) argument for an AI + teacher model of EFL writing feedback. Their study showed that integrating ChatGPT analysis with teacher judgment allowed technological efficiency to be combined with pedagogical interpretation. Similarly, Guo and Wang (2024) emphasized both the opportunities and constraints involved in using ChatGPT to support teacher feedback in EFL writing.

The need for mediation is especially evident when ArguMap-AI misclassifies argumentative components. Automated argument analysis is probabilistic rather than infallible. Pereira et al. (2025) identify linguistic variation, implicit arguments, and contextual interpretation as continuing challenges for educational argument mining, while Schaller et al. (2025) demonstrated that model performance can decline when argumentative texts are incomplete. Thus, the occasional classification errors observed in this study should not be treated only as technical defects. They also reveal an important pedagogical principle: learners need sufficient AI literacy to understand that automated classification is an interpretation requiring human verification.

Technological and Pedagogical Implications

The findings suggest several priorities for future ArguMap-AI development. Technologically, the system needs to improve its ability to distinguish among claim, evidence, reasoning, counterargument, and rebuttal and to analyze not merely the presence of these components but also their relationships. This distinction is important because an essay may contain both claims and evidence while still demonstrating weak reasoning. *Second*, the system should move toward adaptive diagnostic feedback. Rather than providing the same type of warning for every weak relationship, feedback could vary according to the detected problem. For example:

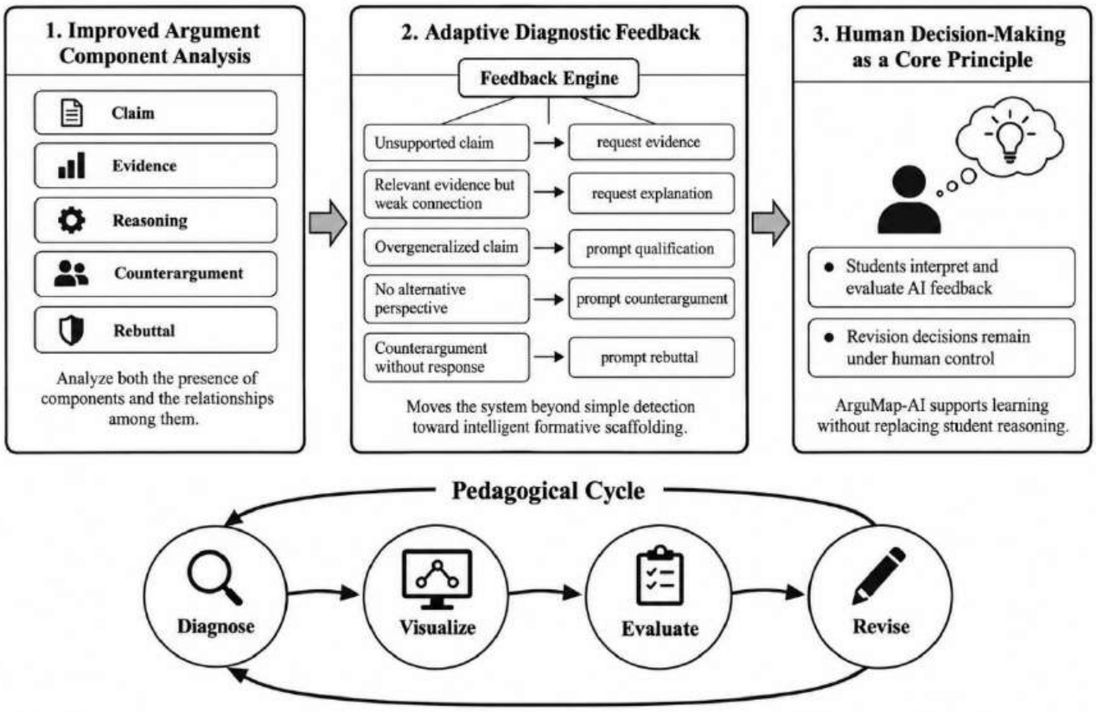

Figure 2. Technological and pedagogical implications for the future development of ArgeMap-AI

Such differentiation would move ArguMap-AI beyond argument-component detection toward intelligent formative scaffolding. Third, the system should retain human decision-making as a core design principle. Automated feedback research increasingly demonstrates that effective AI-supported writing requires learners to engage with feedback rather than simply receive it. Guo et al. (2024) found positive effects of AI support on feedback provision and writing ability, while Weng et al. (2024) showed that recognizing feedback quality does not automatically guarantee effective uptake. From a pedagogical perspective, ArguMap-AI can therefore be conceptualized as part of a diagnose–visualize–evaluate–revise cycle rather than as an automated writing service.

Limitations and Future Research

Several limitations should qualify the interpretation of the findings. First, the study involved only 21 fifth-semester students from one Writing Project course in the English Education Study Program at Universitas Muhammadiyah Enrekang. The findings therefore represent a specific EFL higher-education context and should not be generalized without further investigation.

Second, improvements between initial and revised drafts cannot be attributed exclusively to ArguMap-AI. Students also received lecturer guidance and accumulated experience with argumentative writing during implementation. Comparative studies involving control or alternative-feedback groups would provide stronger evidence regarding the specific contribution of intelligent claim–evidence mapping.

Third, the system's classifications were not always accurate, particularly when argumentative relationships were implicit or linguistically complex. This limitation is consistent with broader argument-mining research showing that incomplete and evolving student texts remain technically challenging (Schaller et al., 2025). Future work should therefore evaluate the system using precision, recall, F1-score, and human-expert agreement in addition to classroom outcomes.

Future studies should also examine ArguMap-AI with larger and more diverse EFL populations and compare text-only AI feedback, visual argument mapping, teacher feedback, and combined human–AI feedback. Longitudinal research would be particularly valuable for determining whether repeated use develops students' independent capacity to identify claim–evidence relationships without technological assistance.

Overall, the findings suggest that the primary contribution of ArguMap-AI is not simply that it provides another source of automated feedback. Its more distinctive contribution is the attempt to externalize argumentative reasoning and return it to the learner for evaluation. In this model, intelligent technology does not complete the student's argumentative work; it makes weaknesses in that work more visible so that the student can reconsider, justify, and revise the argument.

CONCLUSION

This study developed and evaluated ArguMap-AI, an intelligent claim–evidence mapping system for fifth-semester students in the Writing Project course of the English Education Study Program at Universitas Muhammadiyah Enrekang. The findings indicate that students' argumentative writing improved across the assessed dimensions, with the most substantial changes occurring in claim–evidence alignment, reasoning and explanation, and evidence relevance and sufficiency. Students generally perceived visual mapping and diagnostic feedback as useful for identifying unsupported claims and reconsidering the structure of their arguments. Nevertheless, occasional component misclassification, increasingly complex maps for longer texts, and differences in students' ability to interpret AI feedback demonstrate that intelligent writing support continues to require lecturer mediation and critical learner judgment.

Conceptually, ArguMap-AI contributes to AI-assisted EFL writing by repositioning artificial intelligence from a text generator toward a reasoning scaffold. Its proposed functional design can be represented as Writing Input → Argument Component Identification → Claim–Evidence Relationship Analysis → Visual Argument Mapping → Diagnostic Feedback → Student Evaluation → Revision → Reanalysis. Future development may extend this architecture through an argument-classification engine, claim–evidence alignment scoring, adaptive feedback generation, multilevel visualization, and revision-tracking mechanisms. Such integration provides a foundation for developing ArguMap-AI as a computer-implemented educational technology invention, although any future patent application would require separate prior-art searching, technical validation, and assessment of novelty and inventive step before detailed technical disclosure.

The present study remains limited by its small single-class sample and the developmental nature of the prototype. Future research should evaluate ArguMap-AI across institutions and proficiency levels, assess the technical accuracy of its argumentative classifications, and compare it with text-based GenAI feedback and teacher feedback. Further development should preserve a human-in-the-loop principle in which AI diagnoses and visualizes argumentative structure while learners remain responsible for evidence selection, reasoning, and revision. In this form, ArguMap-AI has the potential to function not merely as an automated writing application but as an intelligent environment for cultivating more visible, evidence-based, and independently constructed academic reasoning.

Acknowledgment
The authors gratefully acknowledge the support of Universitas Muhammadiyah Enrekang and thank the students who participated in this study.

Conflict of Interest
The authors declare that they have no conflict of interest related to this study.

Author Contributions
Dr. Mustakim, M.Pd.: Conceptualization, system development, data collection, data analysis, and manuscript drafting. **Dr. Ismail, M.Pd. (Corresponding Author):** Methodology, supervision, data interpretation, manuscript review, editing, and correspondence with the journal. **Sri Rosmiana, M.Pd.:** Literature review, data organization, interpretation of findings, and manuscript revision. All authors read and approved the final manuscript.

Generative AI Disclosure
Generative AI tools were used to support language refinement, structural organization, and editorial improvement of the manuscript. All AI-assisted content was critically reviewed, verified, and revised by the authors, who remain fully responsible for the accuracy, originality, and integrity of the work.

REFERENCES

[1] Alnemrat, A., Aldamen, H., Almashour, M., Al-Deaibes, M., & AlSharefeen, R. (2025). AI vs. teacher feedback on EFL argumentative writing: A quantitative study. *Frontiers in Education, 10*, 1614673. DOI: 10.3389/feduc.2025.1614673.

[2] Alsmari, N. A. (2022). Integrating computer-aided argument mapping into EFL learners' argumentative writing: Evidence from Saudi Arabia. *International Journal of Advanced Computer Science and Applications, 13*(10), 98–104. DOI: 10.14569/IJACSA.2022.0131013.

[3] Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing, 57*, 100745. https://doi.org/10.1016/j.asw.2023.100745

[4] Cengiz, B. C., Bilki, Z., Ataş, A. H., & Çelik, B. (2025). Exploring second language writers' engagement with ChatGPT feedback: Revision behaviors and perceptions. *System, 134*, 103837. DOI: 10.1016/j.system.2025.103837.

[5] Chen, X., Jia, B., Peng, X., Zhao, H., Yao, J., Wang, Z., & Zhu, S. (2025). Effects of ChatGPT and argument map (AM)-supported online argumentation on college students' critical thinking skills and perceptions. *Education and Information Technologies, 30*(12), 17623–17658. DOI: 10.1007/s10639-025-13471-2.

[6] Dewi, U., Daulay, S. H., & Ismahani, S. (2026). AI feedback vs. peer feedback in EFL writing: Exploring Indonesian students' perceptions. *Asian-Pacific Journal of Second and Foreign Language Education, 11*, 49. DOI: 10.1186/s40862-026-00424-6.

[7] Ding, Y., Wehrhahn, F., & Horbach, A. (2025). FEAT-writing: An interactive training system for argumentative writing. In *Proceedings of the 31st International Conference on Computational Linguistics: System Demonstrations* (pp. 217–225). Association for Computational Linguistics.

[8] Gu, Y., Mei, B., & Jing, G. (2026). Enhancing EFL students' argumentative writing with collaborative argument mapping in the GenAI era. *RELC Journal*. Advance online publication. DOI: 10.1177/00336882261471566.

[9] Guo, K. (2024). EvaluMate: Using AI to support students' feedback provision in peer assessment for writing. *Assessing Writing, 61*, 100864. https://doi.org/10.1016/j.asw.2024.100864

[10] Guo, K., & Li, D. (2024). Understanding EFL students' use of self-made AI chatbots as personalized writing assistance tools: A mixed methods study. *System, 124*, 103362. https://doi.org/10.1016/j.system.2024.103362

[11] Guo, K., Li, Y., Li, Y., & Chu, S. K. W. (2024). Understanding EFL students' chatbot-assisted argumentative writing: An activity theory perspective. *Education and Information Technologies, 29*. https://doi.org/10.1007/s10639-023-12230-5

[12] Guo, K., Pan, M., Li, Y., & Lai, C. (2024). Effects of an AI-supported approach to peer feedback on university EFL students' feedback quality and writing ability. *The Internet and Higher Education, 63*, 100962. https://doi.org/10.1016/j.iheduc.2024.100962

[13] Guo, K., Wang, J., & Chu, S. K. W. (2022). Using chatbots to scaffold EFL students'

argumentative writing. *Assessing Writing, 54*, 100666. https://doi.org/10.1016/j.asw.2022.100666

[14] Han, J., & Li, M. (2024). Exploring ChatGPT-supported teacher feedback in the EFL context. *System, 126*, 103502. DOI: 10.1016/j.system.2024.103502.

[15] Liu, J., Shi, Z., & Li, W. (2026). Developing students' feedback literacy in disciplinary academic writing through generative artificial intelligence. *Assessing Writing, 68*, 101030. DOI: 10.1016/j.asw.2026.101030.

[16] Mekheimer, M. (2025). Generative AI-assisted feedback and EFL writing: A study on proficiency, revision frequency and writing quality. *Discover Education, 4*, 170. DOI: 10.1007/s44217-025-00602-7.

[17] Mujeeb, N., Lodhi, M. A., Ali, A., Mashwani, W. K., Mostafa, S. M., & Ahmadini, A. A. H. (2026). Beyond red ink: AI feedback and ESL academic writing development in higher education. *Language Testing in Asia, 16*, 24. DOI: 10.1186/s40468-026-00442-8.

[18] Nesbit, J. C., & Liu, Q. (2025). Argument mapping in higher education: A systematic review. *Higher Education Quarterly, 79*(4), e70063. DOI: 10.1111/hequ.70063.

[19] Ngo, T. T.-N., Chen, H. H.-J., & Lai, K. K.-W. (2024). The effectiveness of automated writing evaluation in EFL/ESL writing: A three-level meta-analysis. *Interactive Learning Environments, 32*(2), 727–744. DOI: 10.1080/10494820.2022.2096642.

[20] Pereira, D. E., Gomes, D. T. S., Vasconcelos, L. L., & Campelo, C. E. C. (2025). A comprehensive survey of argument mining in the educational domain: Techniques, applications, and future directions. *WIREs Data Mining and Knowledge Discovery, 15*(3), e70041. DOI: 10.1002/widm.70041.

[21] Pratama, R. D., Widiati, U., & Hakim, L. N. (2025). Effects of human-AI collaborative writing strategy on EFL students' argumentative writing skills. *Computer Assisted Language Learning*, 1–32. https://doi.org/10.1080/09588221.2025.2503900

[22] Robillos, R. J., & Thongpai, J. (2022). Computer-aided argument mapping within metacognitive approach: Its impact on students' argumentative writing performance and self-regulated learning. *LEARN Journal: Language Education and Acquisition Research Network, 15*(2), 160–186.

[23] Schaller, N.-J., Ding, Y., Jansen, T., & Horbach, A. (2025). Don't score too early! Evaluating argument mining models on incomplete essays. In *Proceedings of the 20th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2025)* (pp. 345–355). Association for Computational Linguistics. DOI: 10.18653/v1/2025.bea-1.27.

[24] Weng, F., Zhao, C. G., & Chen, S. (2024). Effects of peer feedback in English writing classes on EFL students' writing feedback literacy. *Assessing Writing, 61*, 100874. https://doi.org/10.1016/j.asw.2024.100874

[25] Woo, D. J., Wang, D., Guo, K., & Susanto, H. (2024). Teaching EFL students to write with ChatGPT: Students' motivation to learn, cognitive load, and satisfaction with the learning process. *Education and Information Technologies, 29*, 24963–24990. https://doi.org/10.1007/s10639-024-12819-4

[26] Zhai, N., & Ma, X. (2023). The effectiveness of automated writing evaluation on writing quality: A meta-analysis. *Journal of Educational Computing Research, 61*(4), 875–900. DOI: 10.1177/07356331221127300.

[27] Zou, B., Wang, C., He, H., Li, C., Purwanto, E., & Wang, P. (2025). Enhancing EFL writing with visualised GenAI feedback: A cognitive affective theory of learning perspective on revision quality, emotional response, and human-computer interaction. *Learning and Motivation, 91*, 102158. https://doi.org/10.1016/j.lmot.2025.102158