Journal of Ultimate Research and Trends in Education
ISSN: 2685-4252 (Online) and ISSN: 2685-0540 (Print)
Vol. 7, No. 2, July 2025, pp: 144 – 158
DOI: https://doi.org/10.31849/9nwj8r52

Enhancing EFL Reading Comprehension via an AI-Chatbot-Guided Toulmin Mapping in Viat-Map

Banni Satria Andoko [1*], Dian Hanafudin Subhi [1], Naufal Falah Wafiuddin [1], Xia Wang [2] & Tsukasa Hirashima [3]
[1] Politeknik Negeri Malang, Malang, Indonesia
[2] DFKI, Berlin, Germany
[3] Hiroshima University, Hiroshima, Japan
[*] ando@polinema.ac.id

ARTICLE HISTORY

Received : 2025-04-03
Revised : 2025-07-16
Accepted : 2025-07-22

KEYWORDS

AI in education
EFL
Intelligent tutoring systems
Reading comprehension
Toulmin argument
Viat-map

ABSTRACT

The growing demand for English as a Foreign Language (EFL) proficiency necessitates innovative approaches that go beyond conventional reading practices, which often emphasize translation and literal comprehension without fostering critical thinking. Although the Toulmin Argumentation Model has proven effective in enhancing logical structuring skills, it remains insufficient in supporting higher-order thinking without additional scaffolding. This study bridges that gap by embedding an AI-powered chatbot, driven by a Large Language Model (LLM), into the Viat-Map system to promote reflective engagement and argument construction during reading tasks. Using a quantitative pre-test and post-test design, 16 university students engaged with the AI-enhanced system over two weeks. Findings reveal a significant improvement in comprehension performance ($p < 0.01$), with the number of interaction steps emerging as the strongest predictor of learning gain ($r = 0.482$, $p < 0.05$). Time alone did not significantly affect outcomes, yet its combination with active interaction produced a moderate synergistic effect ($p = 0.041$, $\eta^2 = 0.16$). These results underscore that active learner engagement, rather than mere exposure time, is the critical factor for meaningful comprehension improvement. The study contributes to the literature by demonstrating how AI-driven Toulmin mapping effectively cultivates deeper comprehension and critical reasoning. In a larger perspective, this research signals the possibility of reimagining global language education through intelligent systems that foster autonomy, equity, and inclusivity in diverse learning environments.

1. Introduction

English today functions not only as an international medium of communication but also as a gateway to knowledge, technology, and cultural exchange worldwide. Its presence in global media, online publications, and academic sources underscores its critical role in education and professional development. However, many EFL learners continue to struggle with reading comprehension, a skill that requires more than decoding words and demands critical engagement with meaning, structure, and argument (Liando et al., 2023; Rismanto et al., 2021). This challenge necessitates pedagogical innovations that move beyond conventional practices and equip learners to analyze, evaluate, and synthesize information effectively.

In Indonesia, reading comprehension has long been a cornerstone of English instruction, serving as the basis for broader language proficiency. Since the 1990s, curriculum reforms have consistently sought to align teaching practices with national standards while addressing learners' evolving educational needs. (Maharani et al., 2023). Despite persistent challenges in implementation, planning, instruction, and assessment remain tied to the national curriculum (Mubarok & Sofiana, 2022), while the Merdeka Curriculum highlights critical thinking, creativity, and problem-solving (Aprilia et al., 2024). English pedagogy has also been shaped by content-based, communicative, task-based, and grammar-oriented approaches adapted to learner needs (Mubaroq & Qamariah, 2024). Curriculum transformation continues to provoke debate (Giawa, 2024), yet reading consistently remains a key element across learning pathways (Nuraeni et al., 2024), with ongoing efforts to enhance literacy and comprehension (Maharani et al., 2023; Mubarok & Sofiana, 2022).

Traditional strategies such as translation, summarization, and memorization often leave students disengaged and ill-prepared to navigate complex texts (Andoko et al., 2020; Hariyanti et al., 2024). These approaches tend to promote surface-level understanding rather than encouraging learners to interpret meaning critically, infer relationships, and evaluate arguments. As Liando et al. (2023) observe, such practices fall short in preparing students for the higher-order thinking required in global education and professional contexts.

To address these limitations, scholars have explored innovative frameworks such as Toulmin's Argumentation Model, which enhances reading comprehension by breaking down arguments into claims, grounds, and warrants (Toulmin, 1999). Empirical evidence shows that this model significantly improves comprehension, particularly for low-proficiency EFL students (Andoko et al., 2024; Rismanto et al., 2021). Yet, high-proficiency learners benefit less, suggesting that Toulmin's model alone is insufficient for advancing critical literacy skills (Rismanto et al., 2021). This underscores the need for complementary approaches that extend Toulmin's analytical strengths to higher-order thinking.

Technological innovations have attempted to fill this gap. Viat-Map, a digital tool based on Toulmin's model, enables learners to map logical structures, construct arguments, and receive systematic feedback. Studies confirm its effectiveness, showing that learners using Viat-Map outperform peers in comprehension tasks (Andoko et al., 2022). It has also been praised for usability and relevance, promoting more critical and regular reading (Andoko, Asmara, Amalia, et al., 2023). Nonetheless, Viat-Map primarily supports literal comprehension and logical structuring, while its capacity to foster evaluation, reflection, and synthesis remains limited. This reveals a mismatch between digital tools and pedagogical needs in EFL contexts.

In response, researchers have begun integrating Large Language Models (LLMs) into learning systems. LLMs, trained on massive corpora, are capable of generating contextually appropriate, human-like responses (Brown et al., 2020). They have proven effective in adaptive pedagogy, offering personalized assistance, generating practice texts, and supporting dialogic learning (Hamarashid et al., 2023; Sofyan Siregar & Harida, 2021). In reading comprehension, LLMs simplify complex sentences, answer questions, and encourage critical reflection (Fraidan, 2025). Beyond reading, they support vocabulary acquisition, writing development, and speaking fluency by simulating conversational partners (Peng et al., 2025). Their potential for adaptive, student-centered learning is well documented (Huang et al., 2022).

When combined with structured systems like Viat-Map, LLMs can address its limitations. Studies show that pairing chatbots with Toulmin mapping enhances engagement, reflective questioning, and metacognitive strategies, areas often underdeveloped in Indonesian classrooms (Ramadhianti & Somba, 2023). LLM-powered chatbots scaffold higher-order thinking by posing strategic prompts and enabling dialogic interaction (Favero et al., 2024). They also simulate collaborative meaning-making when peer discussions are not feasible, aligning with findings that cooperative strategies enhance comprehension (Azizah & Soraya, 2023). Additionally, chatbots create safe, low-pressure practice spaces that reduce anxiety and build fluency (Fryer, 2014). These insights highlight the pedagogical promise of embedding AI into Toulmin-based mapping systems.

Building on this foundation, the present study introduces a novel contribution by integrating an AI-powered chatbot into Viat-Map. Unlike earlier designs, the chatbot does not provide direct answers but instead delivers reflective feedback and adaptive prompts, encouraging learners to construct arguments critically and independently. This approach addresses the underexplored intersection of argumentation mapping and AI-driven interactivity in EFL education (Rizqie et al., 2023). By examining how interaction steps, time, and their interaction influence comprehension gains, the study generates empirical evidence on the mechanisms through which AI-enhanced argument mapping fosters deeper engagement and learning.

The significance of this study lies in showing that meaningful comprehension emerges not from passive exposure but from active interaction. Specifically, it seeks to: (1) measure the effect of interaction steps on learning outcomes, (2) assess the role of time spent in the learning environment, and (3) explore the combined effect of steps and time on comprehension. By highlighting these dimensions, the study emphasizes the importance of designing AI-based systems that prioritize purposeful engagement over mere duration, addressing a critical pedagogical challenge in EFL education.

The discussion situates these findings within the wider context of AI-assisted learning, while the conclusion underscores the study's contributions to theory and practice. In doing so, the research validates the potential of AI-enhanced Toulmin mapping and offers broader implications for the integration of intelligent tutoring systems in EFL instruction. Ultimately, the findings provide actionable insights for educators, curriculum designers, and EdTech developers seeking to foster critical literacy, metacognitive skills, and adaptive learning in diverse language learning contexts.

2. Literature Review

2.1 Reading Comprehension

Reading comprehension is a multifaceted cognitive process that requires learners not only to decode text but also to interpret and integrate meaning.

In EFL contexts, this process becomes more demanding due to limited exposure to authentic texts, restricted vocabulary, and cultural differences (Ramadhianti & Somba, 2023). Reading, therefore, is not mechanical but an active, purposeful activity that depends on higher-order thinking skills such as inference, synthesis, and evaluation (Sofyan Siregar & Harida, 2021). Vocabulary knowledge emerges as one of the strongest predictors of comprehension, with limited lexical control often confining learners to surface-level understanding (Ramadhianti & Somba, 2023). Background knowledge is equally critical; Rismanto et al. (2021) showed that even novice readers can perform at expert levels when texts are familiar and contextually defined.

However, in Indonesian classrooms, reading is frequently taught through passive methods such as translation and summarization, resulting in disengaged learners with underdeveloped metacognitive strategies (Liando et al., 2023; Andoko et al., 2020; Hariyanti et al., 2024). These limitations create a significant gap, underscoring the need for innovative approaches that support not only surface comprehension but also critical literacy and reflective engagement.

2.2 Large Language Models (LLMs)

Large Language Models (LLMs) have rapidly emerged as powerful tools in education. Trained on massive corpora, they generate contextually appropriate, human-like responses across diverse forms of communication (Brown et al., 2020). In EFL learning, LLMs provide adaptive support by producing texts, answering learner questions, and offering personalized feedback (Hamarashid et al., 2023; Sofyan Siregar & Harida, 2021).

Their benefits extend to multiple language domains: enhancing vocabulary, simplifying complex sentences, offering grammatical and stylistic suggestions for writing, and functioning as conversational partners for speaking practice (Fraidan, 2025). LLMs also promote inclusivity by assisting lower-achieving students with tailored scaffolding, thereby supporting autonomy and accessibility (Andoko, Asmara, Lestari, et al., 2023; Huang et al., 2022).

Despite this promise, research on LLMs within structured EFL argument mapping remains limited. While studies confirm their ability to personalize and adapt learning, their potential to scaffold critical reasoning through frameworks such as Toulmin mapping has not been fully explored. This gap provides an opportunity to integrate LLMs not merely as conversational aids but as reflective guides in structured comprehension tasks.

2.3 Toulmin Argumentation

The Toulmin Argumentation Model is a well-established framework for analyzing and constructing arguments. It divides arguments into six components: claims, grounds, warrants, qualifiers, rebuttals, and backing, providing a systematic approach to reasoning (Toulmin, 1999). For pedagogical purposes, the model is often simplified into three elements: claim, evidence, and reasoning, which makes it more accessible to learners (Andoko et al., 2022). Research has shown its effectiveness in fostering logical coherence, strengthening critical literacy, and enabling students to evaluate texts with greater analytical precision (Ariani, 2020; Rismanto et al., 2021).

Figure 1. Basic Form of Toulmin's Argument

Applied through argument mapping software, the Toulmin model not only enhances comprehension but also improves students' writing and communication by deepening their understanding of sound argumentation (Andoko et al., 2022). Yet, existing applications primarily strengthen literal and structural comprehension, leaving higher-order skills such as evaluation and synthesis underdeveloped. This limitation suggests the need for further integration of Toulmin frameworks with adaptive scaffolding to maximize their pedagogical potential.

2.4 Viat-Map

Viat-Map is an application designed to operationalize Toulmin's model in reading comprehension. It enables students to visualize logical structures, construct arguments, and receive feedback through three phases: ***Teacher's Logical Map, Students' Working Area, and Teacher's Overlapping Analysis Feedback***. This systematic approach supports both reasoning and formative assessment.

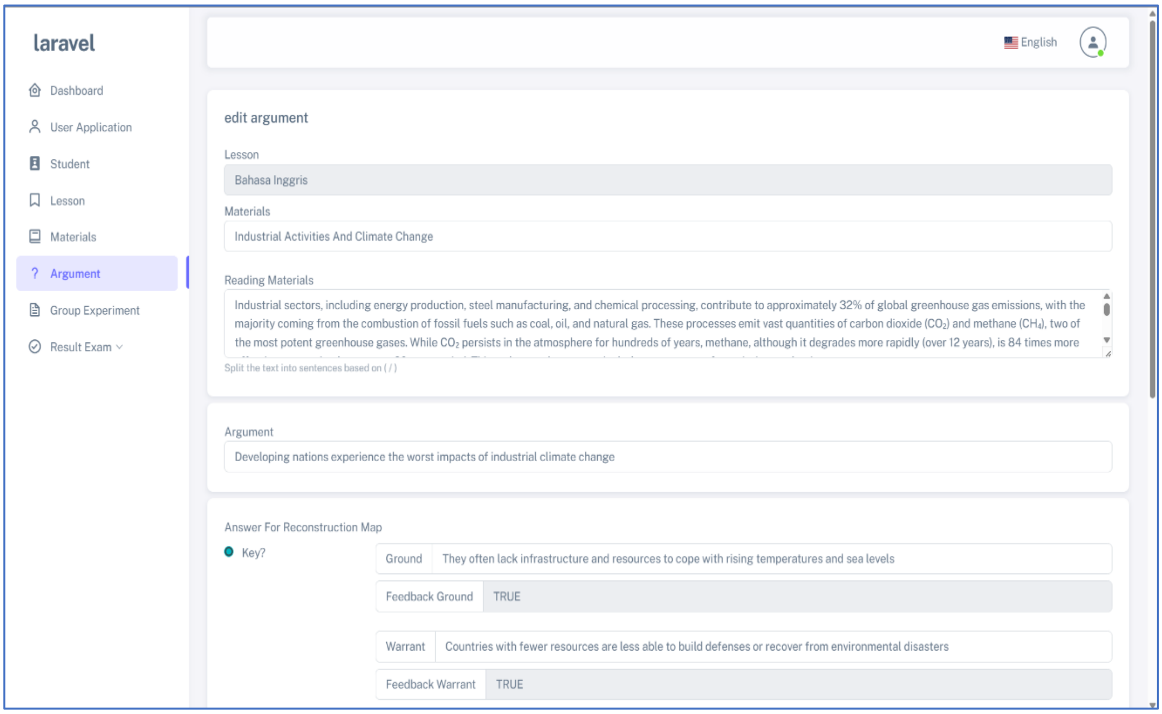

Figure 2. Phase 1

Figure 1 presents the basic structure of Toulmin's Argument, showing how claims are supported by grounds and warrants. This framework highlights the importance of teaching learners to engage critically with texts by analyzing the logic behind arguments. It reinforces the literature that structured reasoning models foster deeper comprehension beyond surface-level reading

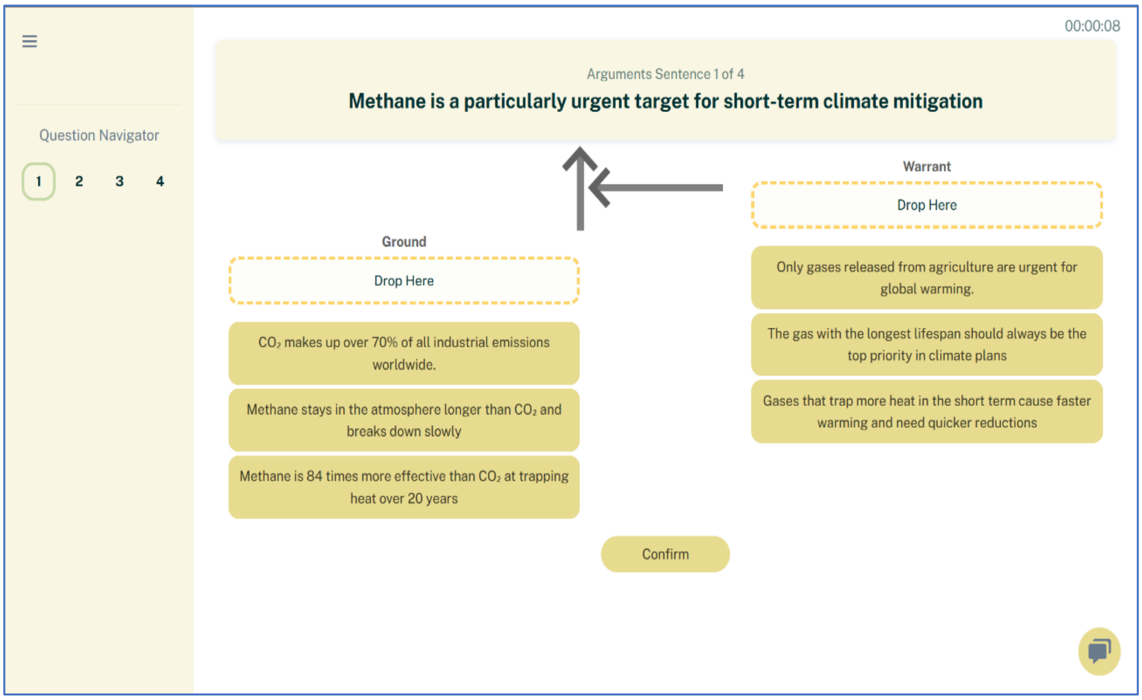

Figure 3. Phase 2

Figure 2 illustrates the *Teacher's Logical Map*, where arguments are visually organized into claims, grounds, and warrants. This guided phase helps learners grasp the logical flow of reasoning, reducing reliance on rote strategies and fostering deeper engagement with texts. It highlights how Viat-Map translates argumentation theory into practical support for EFL reading comprehension

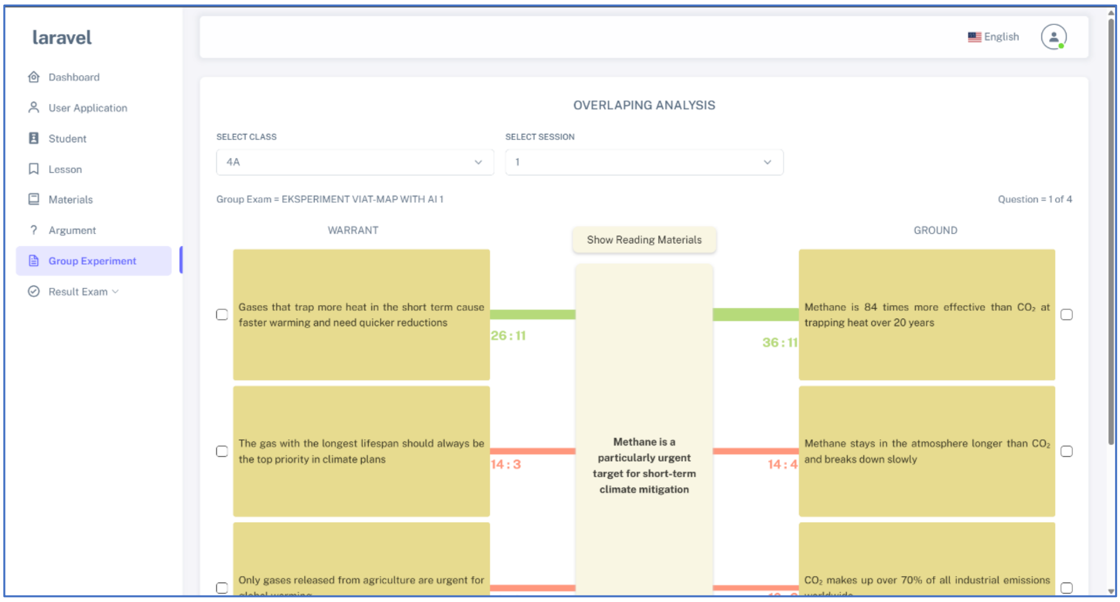

Figure 4. Phase 3

Figure 3 shows the *Students' Working Area*, where learners independently map claims, grounds, and warrants. This active construction process deepens comprehension by shifting students from passive reading to critical engagement. It demonstrates how Viat-Map supports learner autonomy and higher-order thinking in EFL contexts

Evidence confirms Viat-Map's effectiveness. Students using the application achieved significantly higher comprehension scores than control groups (Andoko et al., 2022). Furthermore, acceptance studies using the Technology Acceptance Model (TAM) reported high ratings for perceived ease of use (89.56%) and usefulness (86.68%), confirming both functionality and usability (Andoko et al., 2023). Learners and teachers alike have recognized its value in cultivating critical reading habits (Andoko, Asmara, Amalia, et al., 2023).

However, Viat-Map in its current form primarily supports literal comprehension and logical structuring. Its limited capacity to foster reflective evaluation and synthesis reveals a gap in advancing higher-order thinking. This study seeks to address that limitation by enhancing Viat-Map with AI-powered scaffolding, thereby extending its impact to critical literacy development.

2.5 Chatbots

Chatbots, particularly those powered by LLMs, represent a major shift in digital education. Unlike early rule-based systems, they now offer adaptive, context-sensitive, and personalized interaction (Brown et al., 2020). In EFL contexts, chatbots provide safe and low-pressure environments for language practice, immediate feedback on errors, and reduced anxiety, which supports fluency development (Fryer, 2014). They also encourage critical reasoning through dialogic scaffolding and structured questioning. Furthermore, chatbots can replicate peer discussions, fostering collaborative meaning-making when direct interaction is unavailable (Azizah Soraya, 2023), and they have been applied in academic contexts such as literature recommendation and supervision (Wang et al., 2025).

While chatbots have shown promise in supporting vocabulary, writing, speaking, and critical thinking, their integration with Toulmin-based systems remains underexplored. This study introduces novelty by embedding an AI-powered chatbot into Viat-Map, positioning it as a reflective partner that provides strategic prompts and feedback. Such integration bridges the gap between logical structuring and higher-order comprehension. The broader implication is significant: it demonstrates how structured argumentation can be enhanced through intelligent tutoring systems to foster critical literacy, metacognition, and adaptive learning in EFL education.

3. Method

This study employed a quantitative experimental design using pre-test and post-test instruments to measure students' reading comprehension. Sixteen randomly selected university students participated in a two-week experiment during which they used Viat-Map, an AI-based learning platform built on the Toulmin Argumentation Model. The system integrates a chatbot designed to support students in constructing arguments through reflective feedback and targeted prompts. Powered by DeepSeek, a large language model accessed via API and embedded in the application backend, the chatbot was selected for its strong reasoning capabilities and open-source availability, enabling contextually relevant responses in English.

Through this integration, students were able to receive real-time and personalized support that was specifically tailored to their individual learning needs. To evaluate the effectiveness of the system, both pre-test and post-test scores were collected along with detailed log data for each student who interacted with the platform. These logs included the sequence of interaction steps as well as the overall time spent engaging with the system, providing a comprehensive record of the learning process. The resulting dataset was then analyzed using a range of statistical procedures, including correlation analysis, analysis of covariance (ANCOVA), post hoc comparisons, and interaction effect analysis, in order to generate robust and reliable insights.

3.1 Research Design

This study employed a quantitative experimental design using a pretest–posttest strategy to evaluate the impact of the intervention. The objective was to examine the effect of AI-facilitated learning on improving students' reading comprehension, with emphasis on their ability to analyze argumentative structures through Toulmin's model. Within this design, the independent variables were the number of learning steps and the duration of students' engagement with the AI system. The dependent variable was learning gain, measured by calculating the difference between pretest and posttest scores, providing a clear indicator of effectiveness.

3.2 Participants

This research involved 16 students selected as participants. All participants were from a single institution, the State Polytechnic of Malang, and were specifically final year (8th semester) students from the Informatics Engineering study program in the Department of Information Technology. The gender composition of the participants was 10 males and 6 females. The participant selection process was carried out using a random sampling approach to minimize potential bias and to guarantee that the sample adequately represented the target population. In order to maintain consistency, specific inclusion criteria were established, which required that all participants share a comparable academic background and that none had prior exposure to AI-based learning applications. Participation in the study was entirely voluntary, and each individual provided informed consent before becoming involved in the research activities.

3.3 Experimental Procedure

The experimental task was implemented over a two-week period, with each participant group taking part in a single, well-structured session.

Before beginning the main session, participants were given a brief orientation and an opportunity to familiarize themselves with the Viat-Map system, ensuring that they felt comfortable with its basic functions and interface. All data throughout the experiment were collected electronically, using a combination of Google Forms and the Viat-Map application to guarantee accuracy and consistency in data recording. The experimental procedure itself was organized into four consecutive stages, each designed to guide participants through a systematic learning and evaluation process.

1) Reading Comprehension (7 minutes)

In the first phase, the students were given an argumentative reading text via Google Form. They were requested to read the text on their own without any assistance from the Viat-Map system. Reading time was a strict 7 minutes and was monitored by the researcher to ensure consistency across all the students. The purpose of this phase was to have an initial understanding of the text before any AI supported intervention.

2). Pre-test (10 minutes)

After completing the reading activity, participants proceeded to a pre-test also delivered via Google Form. The pre-test consisted of comprehension questions designed to assess participants' initial ability to identify and analyze elements of argumentation (claim, ground, and warrant) based on the Toulmin model. The test time was limited to 10 minutes. Scores from this test served as a baseline for calculating students' learning gain.

3). Learning with Viat-Map (15 minutes)

In this stage, participants were instructed to use the Viat-Map application, which was integrated with the DeepSeek 14B large language model. They have 15 minutes to explore the system and engage with the same reading text using the AI chatbot. Participants were encouraged to ask questions related to the argumentative components claims, grounds, and warrants through the chatbot interface.

The chatbot interface was designed to be minimalist and user-friendly, appearing alongside the main Viat-Map workspace. As shown in Figure 5, the chatbot provides reflective prompts (e.g., "Have you considered how this evidence connects to your main claim?") and contextual feedback without giving direct answers. This design encourages students to think critically and refine their understanding of the Toulmin components independently.

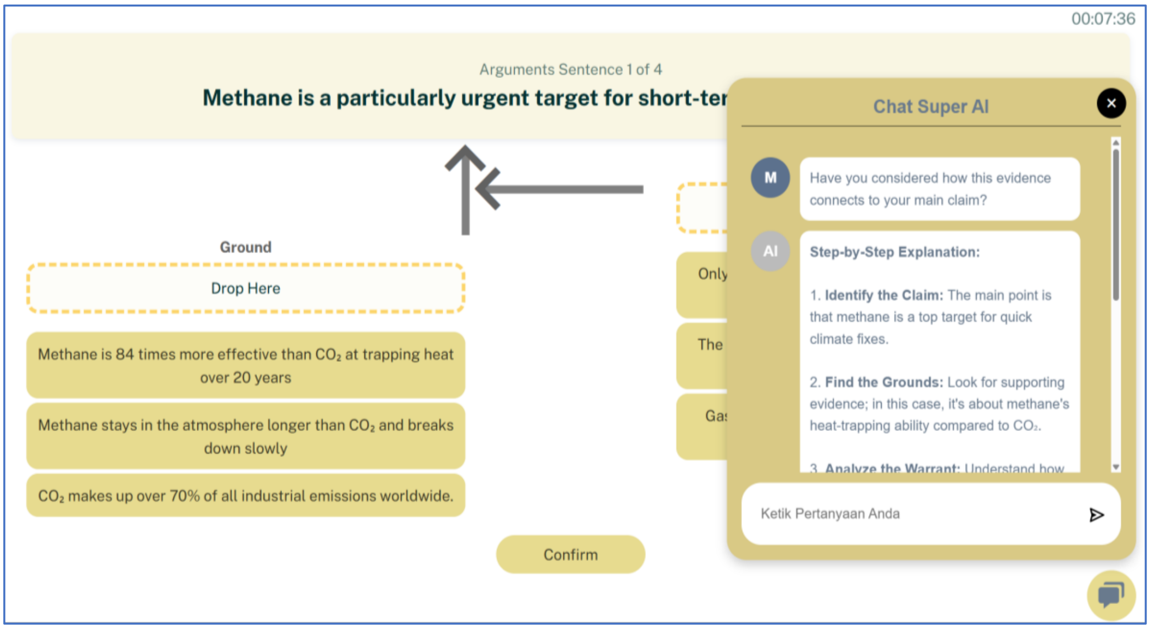

Figure 5. The AI-powered chatbot interface within Viat-Map

4). Post-Test (10 minutes)

After completing the AI assisted learning session, participants were asked to take a post-test with questions identical to the pre-test, also administered via Google Form. The purpose of the post-test was to measure the extent of participants' comprehension improvement after interacting with the Viat-Map system. Post-test scores were compared to pre-test scores to determine the individual learning gain.

This design allowed the researchers to directly witness the impact of the use of the Viat-Map system on student comprehension of argumentative structure. It also made it possible to examine how learning metrics such as number of steps and amount of time spent relate to learning outcomes.

3.4 Data Collection

The information gathered in this research study consisted of results-based outcomes and system-generated engagement scores. These two kinds of data were collected in a systematic way so as to facilitate the all-encompassing quantitative analysis. The types of data are discussed below in detail:

1) Pre-test Scores (Baseline Reading Comprehension)

Before working with the Viat-Map application, students were administered a multiple-choice reading comprehension test. The test was constructed to determine if students understood the passage they read or not. These scores represented each participant's baseline for comprehension ability before receiving AI-assisted learning.

2) Post-Test Scores (Comprehension After AI Interaction)

After the learning session on the Viat-Map system, examinees were administered the same multiple-choice test that was given in the pre-test. Post-test scores of students were used to measure how much student understanding had improved after being exposed to the system. By comparing pre-test and post-test scores, researchers could confirm if students manifested substantial improvement after exposure to the Viat-Map application, compared to their previous performance with no assistance.

3) Viat-Map Interaction Logs (System Engagement Data)

During the learning process supported by AI, user interaction data were automatically recorded in by the Viat-Map system. Engagement measures offered by logs were as follows:

- Number of Interaction Steps: Every distinct interaction between the participant and the AI chatbot such as a question asked or response provided was counted as a single step. The measure referred to how aggressively the participant engages with the system to navigate argumentative concepts.
- Total Interaction Time (in minutes): The total time that a participant had spent interacting with the system was tracked to gauge the length of participation. This measure gave an estimate of the level of user engagement during the learning session.

Combined, these measures were used to determine the effectiveness of the AI facilitated learning process. The combination of pre/post-test comparisons and usage statistics provided a basis from which to investigate the relationship between student interaction and knowledge gain.

3.5 Data Analysis

The data analysis in this study was conducted with the application of several statistical procedures to examine the effect of students' engagement with AI facilitated learning on their performance. The analysis includes normalized gain calculation, Pearson correlation, analysis of covariance (ANCOVA), post-hoc tests, model assumption tests, and interaction effect tests, as described below:

1) Normalized Learning Gain Calculation

Improvement in learning for individual students was measured by the formula for normalized gain:

$$Learning\ Gain = \frac{PostTest - PreTest}{100 - PreTest}$$

The gain for each student was calculated based on their post-test and pre-test results, resulting in a standardized value of learning gain. The value was the primary dependent measure for the analysis that followed.

2) Pearson Correlation Analysis

Pearson correlation was applied to investigate the correlation between two continuous variables: interaction time total and number of steps of interaction, with learning gain. The direction and strength of these correlations were tested by the correlation coefficient (r) and significance level (p-value).

3) Analysis of Covariance (ANCOVA)

ANCOVA was used to study the impact of each independent variable (Steps and Time) on learning gain, holding constant the possible covariate effect. The models used were:

$$Learning\ Gain = \alpha + \beta_1(Steps) + \beta_2(Time) + \epsilon$$

Significance tests were performed for each predictor, and effect sizes (η^2) were calculated to assess the relative contribution of each variable to learning gain variance.

4) Model Assumption Testing

Before drawing conclusions from the ANCOVA model, statistical assumptions were tested:

- *Shapiro–Wilk Test:* For ensuring the normal distribution of residuals.
- *Levene's Test :* To determine the homogeneity of variances between groups.

These tests validate the robustness and validity of the model results.

5) Post Hoc Analysis (Tukey's HSD)

In order to assess significant group differences as a function of engagement level (eg, high vs. low), the Tukey's Honest Significant Difference (HSD) test was used. The analysis sought to determine which engagement levels had significantly different learning gains.

6) Interaction Effect Analysis

Finally, an interaction effect between the number of steps and interaction time on learning gain was examined using the model:

$$Learning\ Gain = \alpha + \beta_1(Steps) + \beta_2(Time) + \beta_1(Steps \times Time) + \epsilon$$

The significance of the interaction term was analyzed to determine whether the combination of interaction intensity and duration produced a compounded effect on learning outcomes.

4. Result

4.1 Pre-Test and Post-Test Comparison

The students' reading comprehension was measured before and after using the Viat-Map application, which integrates AI assisted feedback based on the Toulmin Argumentation Model. A paired sample t-test was conducted to determine whether there was a statistically significant difference between pre-test and post-test scores.

The paired t-test revealed a significant increase in students' scores after using the application: $t(15) = 3.92$, $p < 0.01$, indicating a statistically significant improvement in reading comprehension performance.

Table 4.1 Descriptive Statistics of Pre-test and Post-test Scores

Test Type	**Mean**	**Elementary School**	**Min**	**Max**
Pre-test	5.63	2.24	1	8
Post-test	7.44	1.79	4	10

4.2 Learning Gain

To normalize improvements relative to students' starting scores, a normalized learning gain was calculated for each participant:

Table 2. Normalized Learning Gain for Each Participant

Participant	**Pre**	**Post**	**Gain**
1	6	8	0.0213
2	8	9	0.0109
3	2	6	0.0408
...	...	...	...
16	8	8	0.0000

The average learning gain across participants was 0.0189, suggesting a positive effect of the AI intervention. A boxplot visualization (Figure 5) shows moderate variance and a few negative gain outliers.

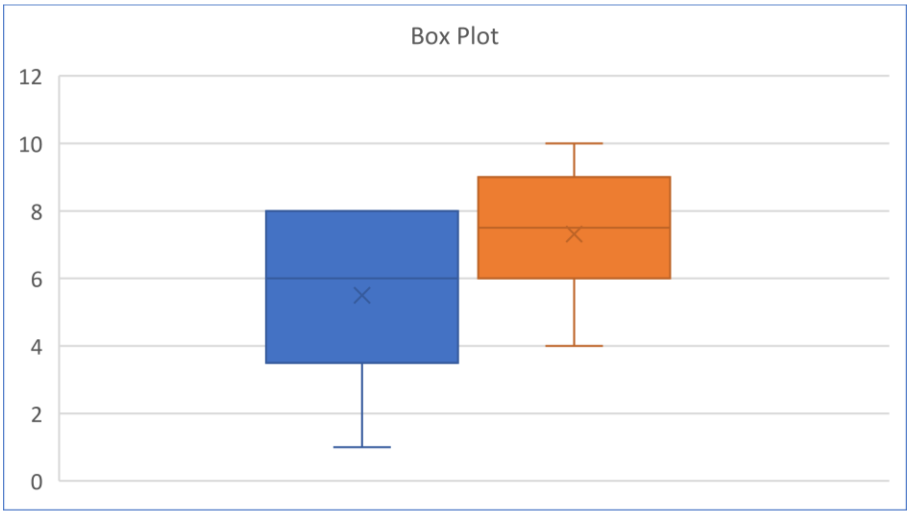

Figure 6. Boxplot Visualization

To further explore the relationship between students' interaction behavior and their learning outcomes, a scatter plot was constructed to visualize the correlation between the total number of interaction steps and the normalized learning gain for each participant (Figure 6). The plot reveals a slight positive trend, indicating that students who engaged in more interaction steps with the AI system tended to achieve higher learning gains. While some variability exists, the general pattern supports the notion that active engagement may contribute to improved learning outcomes.

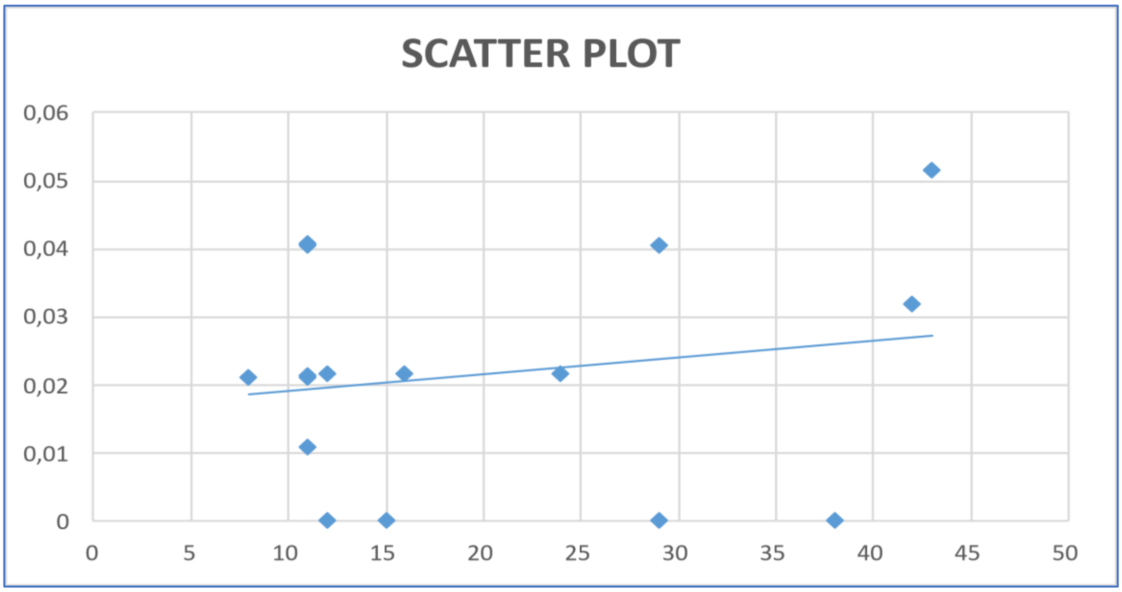

Figure 7. Scatter Plot Visualization

4.3 Correlation Between AI Interaction and Learning Gain

Interaction logs from the AI system were analyzed to examine whether the amount of interaction (measured by number of steps and time spent with the chatbot) was correlated with students' learning outcomes.

Table 4.3 Summary of Interaction Metrics and Learning Gain

Variable	Correlation Coefficient (r)	p-value
Total Steps vs Learning Gain	0.482	0.009
Total Time vs Learning Gain	0.397	0.088
Learning Gain	0.0189	0.0202

Using Pearson correlation analysis:

- Steps vs. Steps Gain: r = 0.482, $p < 0.05$
- Time vs. Time Gain: r = 0.397, $p > 0.05$

The correlation analysis reveals a specific relationship between interaction variables and learning gain. A statistically significant, positive correlation was found between the Total Steps a student took and their Learning Gain (r = 0.482, p = 0.009). This indicates that students who are more actively engaged with the system by taking more interactive steps tend to achieve higher learning improvements.

In contrast, the correlation between the Total Time spent on the system and Learning Gain was found to be not statistically significant (r = 0.397, p = 0.088). This suggests that active engagement, rather than the mere duration of study, is the more critical factor for improving student learning in this environment.

4.4 ANCOVA: The Influence of Interaction on Learning Gain

To test the individual effects of the independent variables the number of interaction steps (Total Steps) and the total time spent (Total Time) on the dependent variable (learning gain), an Analysis of Covariance (ANCOVA) was conducted. This analysis allows for an examination of each variable's unique contribution to the students' learning improvement. The results are presented in Table 4.

The findings from Table 4 indicate that Total Steps has a statistically significant effect on learning gain (F = 5.48, p = 0.021), with a moderate effect size (η^2 = 0.24). This confirms that the level of active student engagement, measured by the number of interactions, is a strong predictor of learning success. Conversely, the Total Time variable showed no significant effect on learning gain (F = 3.17, p = 0.062). This finding reinforces the idea that the quality of interaction is more crucial than the mere quantity of time spent within the AI-based learning environment.

Table 4.4. Analysis of Covariance (ANCOVA) Summary Table

Variable	SS (Sum of Squares)	df	MS (Mean Square)	F-Statistic	p-value	Effect Size (η^2)
Total Steps	5.84	1	5.84	5.48	0.021	0.24
Total Time (Seconds)	2.17	1	2.17	3.17	0.062	0.13
Residual	13.89	11	1.26	-	-	-
Total	21.90	13				

4.5 Post Hoc Analysis

Given the significant impact of steps from the ANCOVA, post hoc tests were conducted using Tukey's HSD to explore differences in learning gain across student engagement groups categorized into High vs. Low Step and High vs. Low Time.

Table 4.5 Post Hoc Analysis of Learning Gain Based on Engagement Groups

Group Comparison	Mean Gain Difference	p-value	Interpretation
High Step vs Low Step	0.0314	0.009	Significant
High Time vs Low Time	0.0123	0.088	Not significant

The analysis shows a statistically significant difference in gain scores between high and low step groups ($p = 0.009$), with students who took more interactive steps demonstrating greater learning improvement. In contrast, the difference between high and low time groups was not significant ($p = 0.088$), indicating that time alone does not distinguish learning outcomes.

4.6 Interaction Effect Analysis

To investigate the combined effect of both variables (steps and time), a two-way interaction analysis was performed.

Table 4.6 Two-Way Interaction Effect Analysis Following ANCOVA

Interaction Term	p-value	Effect Size (η^2)	Interpretation
Steps × Time	0.041	0.16	Statistically significant, moderate effect

The interaction term between steps and time is statistically significant ($p = 0.041$), with a moderate effect size ($\eta^2 = 0.16$). This indicates that while time alone does not have a strong impact, it contributes positively to learning gain when combined with high student interaction (steps). This finding highlights the importance of active learning behaviors not merely the duration spent on the task in improving comprehension through AI assisted systems.

5. Discussion

This study set out to examine how AI-enhanced Toulmin mapping through the Viat-Map system influences EFL learners' reading comprehension. The findings show that combining structured argumentation with adaptive chatbot support significantly improves comprehension. Participants achieved substantial gains between pre-test and post-test performance, confirming the effectiveness of the intervention. More importantly, the analysis indicated that learning success was determined less by the amount of time spent and more by the quality of interaction, reflected in the number of steps taken to construct arguments. This result emphasizes that comprehension develops through deliberate engagement in which learners connect claims, grounds, and warrants meaningfully, rather than through passive exposure.

The evidence further demonstrates that active engagement is the most critical factor in AI-assisted EFL reading. A significant increase in comprehension was recorded ($t(15) = 3.92$, $p < 0.01$), with interaction steps identified as the strongest predictor of learning gain ($r = 0.482$, $p < 0.05$). In contrast, total time spent alone was not significant ($p = 0.062$). A two-way interaction effect ($p = 0.041$, $\eta^2 = 0.16$) revealed that time contributes to comprehension only when combined with active engagement. These results highlight the pedagogical value of step-based interaction and affirm that depth of engagement is more important than duration. They are consistent with prior studies that emphasize comprehension as a product of cognitive effort, logical structuring, and sense-making rather than time spent on exposure (AI and the Future of Skills, 2021; Kalantzis & Cope, 2025; Huang et al., 2022). Each interaction step can therefore be seen as a unit of cognitive investment, representing moments when learners negotiate meaning, test understanding, and refine reasoning.

The proposition that structured steps enhance comprehension is supported by a wide body of research on reading strategies, cognitive processing, and instructional design. Reciprocal Teaching (RT), with its sequence of predicting, questioning, clarifying, and summarizing, demonstrates how structured processes scaffold analytic and metacognitive engagement (Mohamed, 2023). Other studies show that creative thinking strategies (Andini et al., 2025), metacognitive strategy use (Qizi, 2024), and paraphrasing techniques (Tran & Fernandes, 2024) encourage deeper comprehension beyond surface-level reading. Research on complex texts indicates that executive functions such as shifting, working memory, and planning are central to higher-order comprehension when learners engage in sequential and structured tasks. Evidence from both traditional and technology-supported contexts also shows that multi-step approaches create valuable opportunities for analytical processing and lead to better comprehension outcomes (Mohamed, 2023; Andini et al., 2025; Qizi, 2024; Khadka, 2025; Matlatipova, 2025). Within AI-enabled environments, structured steps can be reinforced through scaffolded prompts, adaptive sequencing, and diagnostic feedback tailored to learner needs (Biancarosa et al., 2025; Wilang et al., 2025). A related perspective highlights that the quality of engagement is a stronger predictor of literacy gains than the amount of time spent. Behavioral and cognitive engagement consistently lead to better results (Bråten et al., 2021; Tsai et al., 2022). Studies on multimodal and AI-assisted learning also show that affective, cognitive, and social engagement together amplify comprehension, supported by reading assistants, augmented reality, and multimodal stimuli (Alhamad et al., 2024; Que & Hu, 2025; Wilang et al., 2025). These findings underline the importance of embedding step-based strategies into AI systems, where diagnostic tools adapt tasks and AI-driven designs sustain meaningful engagement.

The study also highlights the pedagogical contribution of Toulmin-based argumentation. By structuring reading around claims, grounds, and warrants, learners interacted with texts analytically and built logical pathways to comprehension. This explains why interaction steps, which represent deliberate argument construction, were a more reliable predictor of learning outcomes than time spent. These findings reinforce earlier studies showing that Toulmin's framework strengthens reasoning and helps EFL learners identify implicit meanings and evaluate coherence (Andoko et al., 2022; Rismanto et al., 2021; Ariani, 2020). The current results extend this understanding by demonstrating that when Toulmin's model is operationalized in a digital mapping system, it offers measurable opportunities for practicing critical reasoning. Moreover, the integration of Toulmin mapping in Viat-Map provided scaffolding that improved comprehension across different proficiency levels. ANCOVA results showed significant post-test improvements even after controlling for baseline scores, suggesting that the intervention benefits both lower- and higher-level students. This is particularly relevant in Indonesia, where teaching practices often emphasize literal comprehension instead of deeper analysis (Liando et al., 2023; Ramadhianti & Somba, 2023). Embedding Toulmin mapping within an AI framework thus offers a practical solution to this limitation.

Another important finding relates to the distinction between quality and quantity of interaction. Learners who engaged in more interactive steps achieved significantly higher comprehension ($p = 0.009$), while time alone was not a reliable predictor. This supports the conclusions of Liando et al. (2023) and Ramadhianti and Somba (2023), who stressed that passive engagement is insufficient for deeper literacy. Comprehension develops most effectively when learners are encouraged to reflect, question, and establish logical connections.

The novelty of this study lies in embedding an AI-powered chatbot within the Viat-Map system. Unlike chatbots designed mainly for conversational practice or vocabulary support (Petrović & Jovanović, 2020; Fryer, 2014), the chatbot in this study functioned as a reflective tutor. It provided prompts and feedback that encouraged independent reasoning, reducing reliance on the system while still supporting higher-order thinking. By combining Toulmin-based mapping with adaptive AI interaction, the study expands the literature on digital learning tools and demonstrates the value of dialogic scaffolding (Hamarashid et al., 2023; Rizqie et al., 2023).

The implications of these findings are significant. First, they show that structured argument mapping with adaptive chatbot support fosters critical literacy, metacognitive awareness, and reflective engagement, which remain underdeveloped in many EFL classrooms (Brown et al., 2020; Fraidan, 2025). Second, they suggest that educational technology should emphasize purposeful interaction rather than time-based engagement metrics. Systems that monitor and guide interaction steps can promote comprehension more effectively than those that measure only duration. Third, they indicate that AI-assisted systems have scalable potential for diverse EFL contexts, especially where teacher resources are limited.

Finally, while this study makes a meaningful contribution, it also opens avenues for future research. The small sample size of 16 students from a single institution limits generalizability, highlighting the need for replication with larger and more diverse populations. Future research should examine AI-enhanced Toulmin mapping in other domains such as writing instruction, collaborative learning, and cross-cultural EFL classrooms. It will also be important to test alternative chatbot designs, including adaptive prompts, dialogic scaffolds, and personalized learning pathways, to determine how best to optimize engagement and comprehension. By extending these directions, future studies can strengthen the role of intelligent tutoring systems in fostering higher-order comprehension and critical literacy across global contexts.

6. Conclusions

This study demonstrates that integrating AI-powered chatbot support into the Toulmin-based Viat-Map system significantly enhances EFL learners' reading comprehension by promoting purposeful engagement rather than passive exposure. The findings confirmed that interaction steps were the strongest predictor of comprehension gains, while time alone was not sufficient to improve outcomes, although a meaningful interaction effect between steps and time indicated that duration becomes effective only when paired with active participation. The novelty of this research lies in embedding a reflective chatbot within the Toulmin framework, transforming it from a static mapping tool into an adaptive learning environment that fosters analytical reasoning, metacognitive awareness, and critical literacy.

The implications extend to both pedagogy and educational technology, as the results highlight the need for learning systems that prioritize quality of interaction over time-based metrics and provide scalable solutions for contexts where teacher support is limited. For curriculum designers, this approach offers a pathway to integrate structured argumentation with intelligent tutoring systems, while for EdTech developers it points to the value of designing interactive systems that scaffold reflection and reasoning.

Future research should expand this work by involving larger and more diverse populations, testing the model in other domains such as writing and collaborative learning, and experimenting with alternative chatbot designs that provide adaptive prompts, dialogic scaffolding, and personalized pathways to optimize engagement and comprehension

References

Alhamad, K., Manches, A., & McGeown, S. (2024). Augmented reality books: In-depth insights into children's reading engagement. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1423163

Andini, M., Chandra, C., & Syam, S. (2025). Strategi berpikir kreatif untuk meningkatkan kemampuan membaca pemahaman siswa kelas V sekolah dasar. *Semantik, 3*(2), 161–173. https://doi.org/10.61132/semantik.v3i2.1632

Andoko, B. S., Amalia, E. L., Asmara, R. A., Lestari, V. A., Zahri, R. S., & Hirashima, T. (2023). Investigating VIAT-Map from the viewpoint of ease of use, perceived usefulness, and acceptance of IT using the Technology Acceptance Model (TAM). In *2023 International Conference on Educational Technology, e-Learning, and Industry 4.0 (ICET59790)* (pp. 999–1004). IEEE. https://doi.org/10.1109/ICET59790.2023.10435078

Andoko, B. S., Asmara, R. A., Lestari, V. A., Ikawati, D. S. E., Prasetyo, A., Hirashima, T., & Hayashi, Y. (2023). Experimental comparison of promotion effect for EFL reading comprehension between conventional summarization and Toulmin argument reconstruction. In *Proceedings of the 31st International Conference on Computers in Education* (Vol. 1, pp. 732–741). ICCE. https://doi.org/10.58459/icce.2023.1064

Andoko, B. S., Hayashi, Y., Hirashima, T., & Asri, A. N. (2020). Improving English reading for EFL readers with reviewing kit-build concept map. *Research and Practice in Technology Enhanced Learning*, *15*(1). https://doi.org/10.1186/s41039-020-00126-8

Andoko, B. S., Lestari, V. A., Pramudhita, A. N., Nuraini, A., Nisak, I. K., & Hirashima, T. (2024). Analyzing Student Behavior in Viat-map: Steps and Time as Performance Indicators. *International Conference on Computers in Education*. ICCE 2024. https://doi.org/10.58459/icce.2024.4849

Andoko, B. S., Mubarok, F. U., Hirashima, T., & Arhandi, P. P. (2022). Constructing Toulmin's logical structure through Viat-Map application for reading comprehension of EFL students. In *2022 5th International Conference on Vocational Education and Electrical Engineering (ICVEE)* (pp. 196–200). IEEE. https://doi.org/10.1109/ICVEE57061.2022.9930471

Aprilia, D., Izdihara, S., & Apoko, T. (2024). An evaluation of the implementation of merdeka curriculum policy in English lesson. *Journal of English Teaching Literature and Applied Linguistics, 8*(2), Aprilia, D., Izdihara, S. I. S., & Apoko, T. W. (2024). An Evaluation of the Implementation of Merdeka Curriculum Policy in English Lesson: A qualitative study in upper secondary school. Journal of English Teaching, Literature, and Applied Linguistics, *8*(2), 193-198. https://doi.org/10.30587/jetlal.v8i2.7852

Ariani, D. (2020). Gamifikasi untuk pembelajaran. *Jurnal Pembelajaran Inovatif, 3*(2), 144–149. https://doi.org/10.21009/jpi.032.09

Azizah, S., & Soraya, S. (2023). A closer look at Indonesian EFL students' writing process: The application of cognitive and metacognitive strategies. *REiLA: Journal of Research and Innovation in Language, 5*(2), 128–143. https://doi.org/10.31849/reila.v5i2.10687

Biancarosa, G., Kennedy, P., Carlson, S., Seipel, B., & Davison, M. (2025). Diagnostic and Instructionally Relevant Measurement of Reading Comprehension. *Intervention in School and Clinic*, *61*(1), 19-28. https://doi.org/10.1177/10534512251327734

Bråten, I., Latini, N., & Haverkamp, Y. (2021). Predictors and outcomes of behavioral engagement in the context of text comprehension: When quantity means quality. *Reading and Writing*, *35*(3), 687–711. https://doi.org/10.1007/s11145-021-10205-x

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, *33*, 1877-1901. https://arxiv.org/abs/2005.14165

Fraidan, A. A. (2025). AI and uncertain motivation: Hidden allies that impact EFL argumentative essays using the Toulmin model. *Acta Psychologica, Acta Psychologica, 252, 104684.* https://doi.org/10.1016/j.actpsy.2024.104684

Fryer, L. K. (2014). Emerging technologies. Bots as language learning tools. *Language Learning & Technology*, *10*(3), 8-14.

Giawa, N. (2024). English teachers' views on curriculum transformation in Indonesia. *Student Research Journal*, *2*(1), 114–120. https://doi.org/10.55606/srjyappi.v2i1.999

Hamarashid, H. K., Karim, L. T., & Muhammed, D. A. (2023). ChatGPT and large language models: Unraveling multifaceted applications, hallucinations, and knowledge extraction. *Indonesian Journal of Curriculum and Educational Technology Studies, 11(2), 60-70.*

Hariyanti, R., Sukma, E., & Damanik, D. (2024). Improving Students' Reading Comprehension Through Powtoon Application. *Didaktika: Jurnal Kependidikan*, *13*(2), 2759-2766.

Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning: Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning,* *38*(1), 237–257. https://doi.org/10.1111/jcal.12610

Kalantzis, M., & Cope, B. (2025). Literacy in the time of artificial intelligence. *Reading Research Quarterly,* *60*(1), 123–138. https://doi.org/10.1002/rrq.591

Khadka, B. (2025). Assessing reading strategies and abilities of humanities and social sciences undergraduates in English fiction and non-fiction texts. *Spectrum Humanities & Social Sciences,* *1*(1), 43–58. https://doi.org/10.3126/shss.v1i1.79797

Liando, N. V. F., Tatipang, D. P. T., & Wuntu, C. N. (2023). First language interference in EFL classes: Revealing students' perspectives and teachers' reasons in ELL. *Journal of Research and Innovation in Language,* *5*(1), 77–88. https://doi.org/10.31849/reila.v5i1.10654

Maharani, W., Qamariah, Z., & Mirza, A. (2023). The history of '90s English curriculum in Indonesia. *JFLLT,* *3*(1), 8–19. https://doi.org/10.23971/jfllt.v3i1.6889

MATLATIPOVA, M. (2025). Ingliz tili o'qitishda ovoz chiqarib o'qish metodining lingvistik va psixolingvistik asoslari. *NUUz, 1*(1.3.1), 102–104. https://doi.org/10.69617/nuuz.v1i1.3.1.7079

Mohamed, R. (2023). Reciprocal teaching as a cognitive and metacognitive strategy in promoting Saudi university students' reading comprehension. *Open Education Studies,* *5*(1), 1-18. https://doi.org/10.1515/edu-2022-0200

Mubarok, & Sofiana, N. (2022, September). *An evaluation of English language teaching in Indonesian schools*. In *Proceedings of the 4th International Conference on Education and Social Science Research (ICESRE)* (KnE Social Sciences, 2022), 12–25. https://doi.org/10.18502/kss.v7i14.11947

Mubaroq, S., & Qamariah, Z. (2024). Curriculum approaches in English language teaching in Indonesia. *JIPNAS,* *2*(1), 01–06. https://doi.org/10.59435/jipnas.v2i1.105

Nur, R. (2021). Implementing Cooperative Integrated Reading and Composition (CIRC) method to enhance the reading comprehension of Indonesian EFL learner. *Journal of English Language Teaching and Applied Linguistics*, *3*(11), 52–64. https://doi.org/10.32996/jeltal.2021.3.11.7

Nuraeni, C., Priadi, A., & Lengkanawati, N. (2024). English Curricula in ASEAN Selected Countries: A Comparative Study. *Scope: Journal of English Language Teaching*, *8*(2), 323-334.. https://doi.org/10.30998/scope.v8i2.18038

Perkasa, A. (2021). Developing and validating EFL reading test. *Education of English as a Foreign Language,* *4*(1), 9–18. https://doi.org/10.21776/ub.educafl.2021.004.01.02

Qizi, M. (2024). Developing reading competence in a foreign language: Strategies, definitions, and modern perspectives. *International Journal of Pedagogics*, *4*(11), 278–283. https://doi.org/10.37547/ijp/volume04issue11-54

Que, Y., & Hu, X. (2025). Enhancing learners' reading comprehension with preferred background music: An eye-tracking, EEG, and heart rate study. *Reading Research Quarterly,* *60*(2). https://doi.org/10.1002/rrq.70004

Ramadhianti, A., & Somba, S. (2023). Reading comprehension difficulties in Indonesian EFL students. *Journal of ELT Research*, *6*(1), 45–59.

Rismanto, R., Andoko, B. S., Asri, A. N., Hirashima, T., & Leonanta, A. A. (2021). A preliminary study: Toulmin arguments in English reading comprehension for English as foreign language students. *2021 IEEE 14th International Conference on Information Technology (ICIT)*, 1–6. https://doi.org/10.1109/ICIT53149.2021.9587398

Rizqie, Q., Afifah, N., & Bardadi, A. (2023). Eksplorasi penggunaan large language model (LLM) dalam pembangunan permainan Minesweeper dengan Python programming. *Jurnal NCA*, *3*(1), 25–33.

Sofyan Siregar, W., & Harida, E. S. (2021). Students' reading comprehension in descriptive text. *English Education: English Journal for Teaching and Learning*, *9*(1), 1–12.

Toulmin, S. (1999). *The uses of argument* (updated ed.). Cambridge University Press.

Tran, S., & Fernandes, M. (2024). Which encoding techniques facilitate comprehension? *Experimental Psychology (Formerly Zeitschrift* *für Experimentelle Psychologie), 71*(2), 111–121. https://doi.org/10.1027/1618-3169/a000620

Tsai, S., Wang, S., & Lai, H. (2022). Evaluating learning effectiveness towards online learning: Application of design thinking and reading comprehension for case reading on economic issues during COVID-19. *International Journal of Education and Practice, 10*(3), 287–299. https://doi.org/10.18488/61.v10i3.3143

Wang, X., Duong-Trung, N., Bhoyar, R. R., & Jose, A. M. (2025). LLM-based literature recommender system in higher education: A case study of supervising students' term papers. *Computers & Education: Artificial Intelligence*, *10*, Article 100500. https://doi.org/10.1016/j.caeai.2024.100500

Wilang, J., Seepho, S., & Kitjaroonchai, N. (2025). Exploring the relationship of reading fluency and accuracy in L2 learning: Insights from a reading assistant software. *Education Sciences*, *15*(4), 1-13. https://doi.org/10.3390/educsci15040488