EFL Learners' Perceptions of Using Grok AI for Planning Argumentative Essays

Nguyen Hoang Kham*, To Vu Le Ngan, Le Tran Minh Anh

Foreign Languages Center, Nguyen Tat Thanh University, Ho Chi Minh City, Viet Nam
*nhkham@ntt.edu.vn

Abstract

This study examined how B1-level EFL learners perceived the use of Grok AI as a cognitive tool for planning argumentative essays. By adopting a cross-sectional, post-intervention survey design, the study involved 128 participants who utilized structured prompts – integrating Toulmin's model and the Involvement Load Hypothesis (ILH) – to generate and evaluate AI content under teacher guidance. Learners' perceptions of usefulness, ease of use, and ILH elements were assessed via a 19-item Likert-scale questionnaire. The analysis revealed the learners' highly positive perceptions across all dimensions; data planning was rated highest (*Mean* = 4.14), followed by claim (*Mean* = 4.11) and warrant planning (*Mean* = 4.08). This study provides preliminary insights into the application of Toulmin's argumentative model to using Grok AI, focusing on learner perceptions during pre-writing planning. Initial pedagogical considerations for integrating AI into EFL writing instruction are discussed based on these findings.

® 2026 Journal of Science and Technology - NTTU

Received 07/05/2026
Accepted 06/07/2026
Published 28/10/2026

Keywords

EFL; argumentation; Involvement Load Hypothesis; Grok AI.

1 Introduction

1.1 Background of the study

Planning is one of the most crucial stages in the composing process. Within problem-solving accounts of composing, writers who put deliberate effort in pre-writing planning are expected to produce texts that are more structurally coherent and more rhetorically effective than those produced without planning [1]. For learners of English as a Foreign Language (EFL), this role of planning is even more important when learner writers have difficulty in handling multiple tasks at the same time, that is, learning and producing content, learning to write about it, and learning to use a new language both to learn and create the content and to write about it, a concern that runs throughout the research agenda on writing and language learning [1]. Meanwhile, planning before drafting is considered to enable writers to make more effective use of their attentional resources [2]. As such, denying the writer planning time is thought to impede the complexity, accuracy, and/or fluency of language production [3].

Within the broad domain of writing, the argumentative essay places particularly heavy demands on EFL learners during the planning stage [1]. Successful argumentation requires formulating a central claim that is both contestable and defensible; assembling data that adequately and appropriately supports that claim; and constructing the logical relationship between evidence and conclusion [4]. All three moves need to be managed during planning if the resulting draft is to

https://doi.org/10.55401/1jybvb97

possess argumentative integrity, echoing the insistence that an argument fails where one of these core elements is absent or inadequately justified [4].

The recent emergence of large language model (LLM)-powered AI chatbots has introduced a new planning tool for language learners. Chatbot technology has the capacity to respond dynamically to learner input, provide practice opportunities, and support skills development in real time [5]. When a chatbot is structured around a rhetorical framework, such as Toulmin's model, it can guide learners through argumentative planning in a principled, sequenced way while remaining accessible to intermediate-level learners. The Involvement Load Hypothesis (ILH) provides a theoretical framework for examining the depth of cognitive engagement that such AI-assisted planning can induce [6]. The ILH proposes that learning depth is a function of three task-induced components: need (metacognitive awareness of a gap), search (active effort to fill that gap), and evaluation (critical appraisal of candidate solutions). In addition, the hypothesis posits that tasks which engage all three components involve deeper processing than tasks that engage less. By applying this framework to AI-assisted argumentation planning, the present study assumed that chatbot interactions structured to activate need, search, and evaluation for each Toulmin component would produce more effective planning than interactions that merely provide content without activating these processes.

In the Vietnamese EFL context, teaching writing skills tends to be problematic as instruction fails to move beyond a product approach, leaving learners to struggle with fixed move-step structures such as the five-paragraph expository essay [7]. Consequently, EFL learner writers frequently struggle with argumentation, often omitting necessary supporting evidence, and failing to strictly follow the fixed move-step structure [7]. Meanwhile, AI chatbots are widely used for writing instruction but EFL students primarily utilize them for passive text generation rather than active engagement with rhetorical planning, with minimal style appropriateness checking [8]. This pattern reveals a critical gap regarding how AI – human interactions can be pedagogically engineered to foster deeper, task-induced processing during pre-writing.

To address this gap, the present study aims to investigate the EFL learners' perceptions of using Grok AI, a large language model chatbot, to plan argumentative essays scaffolded by Toulmin's claim-data-warrant structure. The intervention was designed to instantiate the ILH components at each argumentative move through researcher-designed prompts, combined with teacher-guided evaluation of AI-generated content.

1.2 Toulmin’s Model of Argumentation

Toulmin's model of argumentation was developed as a practical, non-formal framework for evaluating the quality of everyday reasoning [4]. Its six elements, including claim, data (grounds), warrant, backing, qualifier, and rebuttal, describe the functional architecture of a complete, persuasive argument. Driven by pedagogical purposes, the present study focuses on three core elements: (1) the claim is an assertion in response to a problem, (2) the data serves as the basis for making a claim, i.e., facts or observations about the situation under discussion, and (3) the warrant represents the move from data to claim by featuring common sense rules, laws, or scientific principles, or thoughtfully argued definitions [4].

Several studies have examined the effects of scaffolding instruction with the Toulmin model on the quality of learners' writing [9]. Critically for the present study, this body of research also reveals asymmetry in the attention to core components of Toulmin’s model. Students were found to demonstrate difficulties in finding relevant evidence to support their claims, therefore they are unable to combine evidence and reasons effectively [9]. Compared with claim and data, warrants are difficult to identify because they are implicit in nature, and their quality depends on the logical relationship between reasons, evidence, and claims [9] . This could be because the warrant requires a general, hypothetical statement that can act as a link between data and claim [4]. The different degrees of cognitive difficulty from the most accessible to the

most challenging, provide the primary theoretical rationale for expecting differential learner perceptions across the three AI-assisted planning stages examined in this study.

1.3 Involvement Load Hypothesis

ILH was proposed to explain why some tasks for vocabulary learning entail deeper processing and better retention than others [6]. The construct of involvement is a motivational-cognitive one that consists of three basic components: need, search, and evaluation [6]. These components are defined as follows:

- The need component refers to the motivational dimension of involvement. It has two degrees of prominence: moderate and strong. When need is moderate, it is imposed by an external agent. One example is when the teacher asks for students to use a word in a sentence. When need is strong, it is intrinsically motivated, or self-imposed by the learners, for instance, by the decision to look up a word in a dictionary while composing a text.
- The search component is the attempt to find the meaning of an unknown word or to look for the word form to express a concept.
- The evaluation involves comparing a certain word with others in order to decide whether it fits the context. Evaluation is considered to be moderate when one recognizes differences between words (as in a fill-in task with words provided), or differences between several senses of a word in a given context. Meanwhile, evaluation that requires a decision as to how words combine with the new word in an original sentence or text is considered to be strong.

While the ILH was originally formulated in relation to vocabulary acquisition, subsequent scholarship has broadened its explanatory scope in two directions. On the one hand, writing has been considered as demanding higher involvement load, specifically in terms of need, search, and evaluation than reading [10]. The writing tasks may offer deeper processing benefits, leading to more durable vocabulary retention [10]. On the other hand, researchers have begun applying ILH components to technology-enhanced language learning environments. When designing an L2 vocabulary task, the current digital learning mode could be considered and integrated with the search components of ILH within electronic apps to advance with the times [11]. AI chatbots, by virtue of prompting learners to recognize conceptual and lexical gaps, search responses for appropriate content, and evaluate the suitability of that content, can trigger the three ILH components, which could make them theoretically beneficial to deep processing in L2 writing. Therefore, the present study extends this line of research by applying the ILH not to vocabulary acquisition, but to the task of argumentative planning, examining whether AI-assisted planning through Grok differentially instantiates ILH components across the three Toulmin argumentative stages.

1.4 Grok AI

The present study employed Grok AI, an LLM-based conversational agent that not only has the potential for generating informative, contextually relevant responses, but is also free for learners to use. For the purposes of argumentative essay planning, Grok offers several meaningful affordances for the present study. First, it can generate multiple claims, data, and warrant statements on demand, providing learners with not just vocabulary but also ideas for argumentative planning without requiring them to have extensive prior knowledge of the topic in question. This is a particularly useful feature for B1-level learners whose content-area knowledge may be limited. Second, Grok can also provide evaluative queries, providing comparative assessments of competing claims, data, or warrants that operationalize the ILH evaluation component. Third, Grok's conversational interface allows the interaction to be structured through sequential prompts that mirror the claim–data–warrant structure of Toulmin's model, enabling the researcher to design prompts that guide learners through each argumentative component in a principled order. In conclusion, the present study employed Grok with the aim of engaging learners to participate more actively in dialogues with very specific goals designated by the three components of Toulmin’s model.

2 Research Methods

To find how B1-level EFL learners perceive the use Grok AI to plan for argumentative essays at Nguyen

Tat Thanh University, the research questions concern EFL learners' perceptions of using Grok AI to plan for argumentative writing, as follows:

1. How do the EFL learners perceive the use of Grok AI to plan for claim?
2. How do the EFL learners perceive the use of Grok AI to plan for data?
3. How do the EFL learners perceive the use of Grok AI to plan for warrant?

Addressing the research questions, this study employed a cross-sectional, post-intervention survey design, combining a pedagogical intervention and a post-intervention questionnaire. The former consisted of four sequenced sessions of AI-assisted argumentative planning, and the latter measured learner perceptions of the experience. This design would allow participants' perceptions to be gathered after authentic engagement with the tool under investigation, ensuring that responses reflect direct experiential knowledge rather than uninformed speculation.

2.1 Research site

The present study was grounded in the context of the general English program at Nguyen Tat Thanh University. The program, administered by the Foreign Languages Center, comprises six courses spanning CEFR levels A1 to B1, designed to enhance students' English proficiency. Purposeful sampling was employed to select 128 participants (N = 128; 37 male, 91 female) from three classes, who were using the textbook Personal Best B1. This sample was chosen based on the assumption that the B1 level of proficiency would allow for meaningful experience with argumentative writing. This purposeful sampling does not aim to ensure generalization of results. However, in this study, the experience with using Grok for planning argumentative essays was essential to make sure that the informants have the capacity to provide meaningful data on the phenomenon under scrutiny.

2.2 Research tools

2.2.1 The use of Grok AI for planning argumentative essays

The AI tool chosen for the present study was Grok for the participants to interact with Grok scaffolded by ILH components and Toulmin's model of argumentation. Grok AI was prompted to address the three components of the argumentation, i.e. claim, data, and warrant one by one. Within each component, the ILH steps were also integrated. For example, the prompt to trigger needs for claim was worded as *"I need to write an argumentative essay on [topic]. What position can I take?"* This prompt ensured the planning process followed each of the three argumentative elements with three ILH components.

Table 1 Toulmin-based prompts for learners to interact with Grok

Toulmin's component	ILH	Prompts
Claim	Need	*"I need to write an argumentative essay on [topic]. What position can I take?"*
Claim	Search	*"Give me three possible claims for this topic, from different viewpoints."*
Claim	Evaluation	*"Which of these claims is strongest and why?"*
Data	Need	*"My claim is [X]. What evidence or facts support this claim?"*
Data	Search	*"Give me three different types of evidence (statistics, examples, expert opinion) for my claim."*
Data	Evaluation	*"Which piece of evidence is most convincing for the claim '...?' Why?"*
Warrant	Need	*"Explain why the evidence I chose actually proves my claim."*
Warrant	Search	*"Give me two or three different ways to explain why this evidence supports my claim."*
Warrant	Evaluation	*"Which explanation is the most logical and convincing? What makes it a strong warrant?"*

2.2.2 Pedagogical intervention

The intervention was delivered over four-session period within the participants' regular English course. Session 1 lasted approximately 60 minutes to accommodate the introductory content. During this session, the participants were first instructed on Toulmin's components of argumentation based on a model written text. Afterwards, they were familiarized with using Grok AI to plan for argumentative writing. The participants were asked to raise questions for clarifications, so that they would experience with the tool. Sessions 2, 3, and 4 focused on planning, with each lasting around 45 minutes. This aimed to allow the participants to use Grok to plan for the task in question. These three sessions were designed for learners both to draw on their understanding of topic-relevant practice, and deal with new argumentative essay task to prevent content-specific familiarity effects. Table 2 provides an overview of the four sessions.

Table 2 Overview of the four-session intervention

Session	Duration (min)	Task	Procedure
1	60	Introduction to argumentative writing and Toulmin's model (claim, data, warrant); overview of Grok AI interface	Whole-class instruction; modelled examples of each Toulmin's component
2	45	Claim, data and warrant planning – Task 1: "Is it time to ban smartphones in public?"	Researcher-designed prompts → learner interaction with Grok
3	45	Claim, data and warrant planning – Task 2: "Do we need to ban fast-food restaurants?"	Researcher-designed prompts → learner interaction with Grok
4	45	Claim, data and warrant planning – Task 3: "Do we need stores now that we can buy everything online?"	Researcher-designed prompts → learner interaction with Grok

The intervention also included a planner, allowing the participants to engage in the planning for each argumentative component with need, search and evaluation.

2.2.3 The questionnaire

After the intervention, the participants completed a Likert-scale questionnaire designed to measure their perceptions of the AI-assisted planning experience. There were 19 items, each of which was rated on a five-point scale (1 = Strongly Disagree to 5 = Strongly Agree) and were distributed across five subscales. The first two subscales captured general perceptions of the intervention: usefulness and ease of use. Usefulness (Items 1-5) assessed learners' perceptions of the tool's contribution to planning quality, idea generation, and overall utility. Ease of use (Items 6-10) assessed perceived accessibility of the Toulmin framework and navigability of the Grok interface. These subscales provided contextual information about learners' general experience of the intervention.

The remaining three subscales mapped directly onto the Toulmin components, with each structured around the three ILH sub-dimensions of need, search, and evaluation:

Claim Planning – Involvement (Items 11-13): Item 11 assessed need awareness ('Using Grok AI made me aware of a need for claim'); Item 12 assessed search ('Using Grok AI helped me search for claim'); Item 13 assessed evaluation ('Using Grok AI helped me evaluate for claim').

Data Planning – Involvement (Items 14-16): Items 14, 15, and 16 assessed the same three ILH sub-dimensions applied to the data component.

Warrant Planning – Involvement (Items 17-19): Items 17, 18, and 19 assessed need, search, and evaluation for the warrant component.

This parallel structure across the three Toulmin components aligned with the ILH elements, which was intended to allow learners' perception of using Grok to be investigated, whereby to address the three research questions.

2.3 Data Analysis

The perceptions of using Grok to plan for argumentative essays scaffolded by Toulmin's model and ILH was measured by a 19-item Likert-scale questionnaire. The questionnaire was administered digitally immediately after Session 4. This choice was made to ensure the learners' perceptions were based on direct, recent experience of the four-session intervention. The data were then input and analyzed using SPSS version 20. The results were analyzed using descriptive statistical procedure to provide an overview of the data. Descriptive statistics were computed for each item and subscale. Because the analysis was confined to descriptive statistics, no inferential test estimates were computed; the differences between items and subscales reported below are therefore characterized descriptively. The reliability for the overall questionnaire was assessed using Cronbach's alpha, with $\alpha = 0.89$, indicating good internal consistency. Reliability was also acceptable for each of the five subscales (usefulness, $\alpha = 0.75$; ease of use, $\alpha = 0.78$; claim planning, $\alpha = 0.75$; data planning, $\alpha = 0.74$; warrant planning, $\alpha = 0.79$).

3 Results

The present study aimed to investigate EFL learners' perceptions of using Grok for planning argumentative essays based on Toulmin's model and ILH. The data involved 128 (N = 128) responses, subjected to statistical analysis. This section presents the analysis results of the participants' response to the questionnaires to address the three research questions. Table 3 summarized the descriptive statistics for the questionnaire subscales and items.

Table 3 Descriptive statistics for questionnaire subscales and items (N = 128)

Subscale / Item	***M***	***SD***
Usefulness (Items 1-5)	4.06	0.90
Ease of use (Items 6-10)	4.03	0.80
Claim planning (Items 11-13)	4.11	0.74
Item 11 – Need	4.06	0.79
Item 12 – Search	4.13	0.68
Item 13 – Evaluation	4.14	0.74
Data planning (Items 14-16)	4.14	0.68
Item 14 – Need	4.13	0.68
Item 15 – Search	4.13	0.72
Item 16 – Evaluation	4.17	0.64
Warrant planning (Items 17-19)	4.08	0.77
Item 17 – Need	4.15	0.74
Item 18 – Search	4.12	0.75
Item 19 – Evaluation	3.98	0.80

Research question (RQ) 1 asked how EFL learners perceive the use of Grok AI to plan for claims. In response to RQ1, the Claim subscale produced a composite mean of M = 4.11. Examining the three claim items individually, all demonstrated clearly positive profiles. Item 12 (*Using Grok AI helped me search for a claim*) and Item 13 (*Using Grok AI helped me evaluate my claim*) both produced higher means (M = 4.13 and M = 4.14 respectively) than Item 11 (*Using Grok AI made me aware of a need for a claim*; M = 4.06). This pattern suggests that learners perceived the search and evaluation scaffolding functions of Grok AI as slightly more effective than its need-activation function at the claim stage, though the differences were rather small.

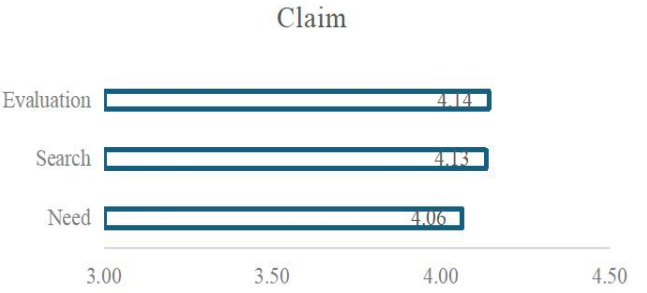


Figure 1 Perceptions of using Grok to plan for claim

Regarding Research question 2, the Data subscale (Items 14-16) produced the highest composite mean among the three Toulmin's components (M = 4.14). At the item level, Item 16 ('*Using Grok AI helped me evaluate my data*') recorded the highest mean in the Data subscale (M = 4.17). Item 14 ('*Using Grok AI made me aware of a need for data*') and Item 15 ('*Using Grok AI helped me search for data'*) both

produced means of M = 4.13. The ratings across all three data planning items suggest that Grok AI was particularly effective in scaffolding for gathering evidence during the argumentative planning.



Figure 2 Perceptions of using Grok to plan for data

Regarding RQ3, the warrant subscale produced a composite mean of M = 4.08, which was the lowest mean among the three RQ subscales. Within the warrant subscale, a notable pattern emerged: Item 17 *('Using Grok AI made me aware of a need for a warrant'*; M = 4.15) and Item 18 *('Using Grok AI helped me search for a warrant'*; M = 4.12) produced slightly higher means than Item 19 ('*Using Grok AI helped me evaluate my warrant'*; M = 3.98). This pattern suggests that while learners found Grok AI effective in raising their awareness of the warrant requirement and in generating warrant content, the evaluative appraisal of warrant quality – the highest-order ILH processing demand – was perceived as less well-supported.

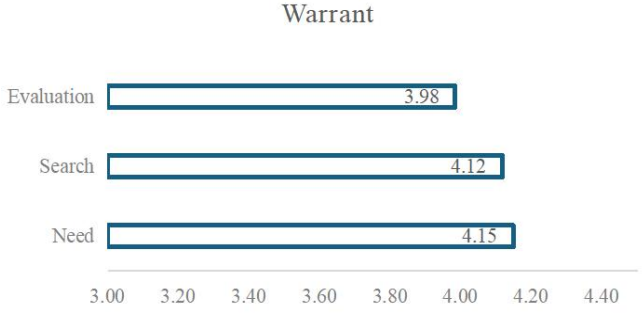


Figure 3 Perceptions of using Grok to plan for warrant

Taken together, the results addressed the three research questions as follows. With respect to RQ1, the learners perceived Grok AI as strongly supportive for claim planning (M = 4.11). Regarding RQ2, data planning generated the highest perceptions of the three RQ subscales (M = 4.14), with Item 16 (data evaluation; M = 4.17) the highest-rated RQ item. As for RQ3, perceptions of warrant planning were slightly lower than those of claim and data (M = 4.08) but remained well above the scale midpoint, with the evaluative dimension (Item 19) the lowest-rated item across all RQ items (M = 3.98). The results indicated that Grok AI was perceived as similarly effective across all three Toulmin planning stages, with only small differences among them. The only exception to this pattern of uniformity was the evaluation sub-dimension of ILH, where warrant evaluation was perceived as lower than claim and data evaluation.

4 Discussion

This study set out to investigate how B1-level EFL learners perceived the use of Grok AI, scaffolded by Toulmin's model of argumentation [4], when planning argumentative essays across three rhetorical components: claim, data, and warrant. Three research questions guided the enquiry, each targeting one Toulmin component and operationalizing learner perception through the three sub-dimensions of the ILH need awareness, search, and evaluation [6]. The statistical analysis yielded three important findings: (1) the learners held positive perceptions of Grok for all the three planning components; (2) the three components were perceived as similarly well-supported, with only small differences among them; (3) one noted pattern emerged whereby warrant evaluation was rated somewhat lower than claim and data evaluation (M = 3.98 vs 4.14 and 4.17), which is consistent with accounts of the warrant as the most demanding component, though the difference is small. There were several limitations inherent to the present study. First, the study measured perceptions only. It examined how learners experienced AI-assisted planning, not what that planning produced: no final argumentative essays were collected, rated or compared. Whether planning with Grok AI actually raises the quality of learners' completed essays, or their course grades, therefore remains an open question, and the positive perceptions reported here should not be read as evidence that it does. Establishing that would require a design incorporating rated final drafts, idealy with a comparison group planning without AI support.

Second, writing argumentative essays served as a site for learning, focusing on the content, ideas, and vocabulary related to the three Toulmin components - claim, data, and warrant - rather than on producing polished final drafts. Third, the analysis was confined to descriptive statistics. Because no inferential tests such as repeated-measures ANOVA were computed, the differences among the subscale and item means reported above are described rather than tested, and none of the comparisons between them can be treated as statistically confirmed.

The finding that claim planning generated strong positive perceptions (M = 4.11) is theoretically consistent with the established literature, which identifies the claim as the most accessible component for learner writers [9]. In EFL contexts, students tend to be heavily trained in fixed move-step structures like the five-paragraph expository essay, where the claim functions as the familiar thesis statement [7]. Against this backdrop, the high perceptions recorded here suggest that Grok AI was able to build productively on this pre-existing familiarity, using it as a foundation for deeper engagement. Data planning generated the highest subscale mean (M = 4.14), and the data evaluation item (Q16; M = 4.17) was the highest-rated item across the questionnaire. This is particularly noteworthy because EFL students frequently demonstrate inadequate argumentative skills characterized by difficulties in finding relevant evidence and a failure to include detailed information such as facts, statistics, or expert opinions [7]. While warrant planning perceptions (M = 4.08) were the lowest of the three Toulmin components, they differed only marginally from claim (M = 4.11) and data (M = 4.14), a gap of at most 0.06 on a 5-point scale. Taken at face value, this finding departs from theoretical predictions that warrant planning is markedly more difficult than claim or data formulation [9]. The literature emphasizes that warrants are implicit in nature, difficult to identify, and often demand more cognitive effort because they depend on the underlying logical relationship between reasons and claims [9]. Therefore, the high level of student perceptions reported here suggests that Grok AI may have functioned as an effective scaffold, helping students externalize the logical links that typically remain unstated or missing in their writing.

5 Conclusion and Recommendations

The present study examined B1-level EFL learners' perceptions of using Grok AI. The findings indicated that planning argumentative essays step by step with Grok AI helped learners engage with Toulmin's elements of argumentation and that they held positive perceptions of Grok's support for all three components, with only small differences among them.

For pedagogy, the findings suggested that teachers can productively integrate Grok-supported Toulmin planning into writing instruction, using claim as an entry point and gradually building learners' confidence in data and warrant planning. For research, future studies could replicate the design with different AI tools or proficiency levels and examine how Grok-supported planning affects both learners' perceptions and the actual quality of their argumentative writing. There are two boundaries of the present study. It documented what learners perceived rather than what they produced, so the effect of Grok-supported planning on essay quality is still unknown; and it reported means and standard deviations without inferential testing, so the differences observed among the three Toulmin components remain descriptive.

For pedagogy, the findings suggested that teachers can productively integrate Grok-supported Toulmin planning into writing instruction, using claim as an entry point and gradually building learners' confidence in data and warrant planning. For research, future studies could replicate the design with different AI tools or proficiency levels and examine how Grok-supported planning affects both learners' perceptions and the actual quality of their argumentative writing.

Acknowledgements

This research is funded by Nguyen Tat Thanh University, Ho Chi Minh City, Viet Nam under grant number 2026.01.134/HĐ-KHCN.

References

1. Manchón, R. M. (2020). Writing and language learning: Looking back and moving forward. In R. M. Manchón (Ed.), *Writing and language learning: Advancing research agendas* (Vol. 56, pp. 3-26). John Benjamins. https://doi.org/10.1075/lllt.56.01man.
2. Baaijen, V. M., & Galbraith, D. (2018). Discovery through writing: Relationships with writing processes and text quality. *Cognition and Instruction*, 36(3), 199-223.
3. Johnson, M. D. (2020). Planning in L1 and L2 writing: Working memory, process, and product-Corrigendum. *Language Teaching*, 53(4), 537-537.
4. Toulmin, S. E. (1958). *The uses of argument*. Cambridge University Press.
5. Fryer, L., Coniam, D., Carpenter, R., & Lăpușneanu, D. (2020). Bots for language learning now: Current and future directions. *Language Learning & Technology*, 24(2), 8-22.
6. Hazrat, M., & Read, J. (2022). Enhancing the Involvement Load Hypothesis as a tool for classroom vocabulary research. *TESOL Quarterly*, 56(1), 387-400. https://doi.org/10.1002/tesq.3051.
7. Pham, V. P. H., & Bui, T. K. L. (2021). Genre-based approach to writing in EFL contexts. *World Journal of English Language*, 11(2), 95-106.
8. Do, H. M. L., Nguyen, T. T. N., Nguyen, L. M. T., Nguyen, L. T. L., & Cao, T. X. L. (2026). Using AI chatbots for self-regulated writing: EFL students' practice, perceptions, and support needs. *Journal of Inquiry into Languages and Cultures*, 10(1), 1-17. https://doi.org/10.63506/jilc.1001.414.
9. Yang, R. (2022). An empirical study on the scaffolding Chinese university students' English argumentative writing based on Toulmin model. *Heliyon*, 8(12), e12199. https://doi.org/10.1016/j.heliyon.2022.e12199 (Corrigendum published in Heliyon, 9(11), e22032).
10. Qi, Y. (2025). A comparative study of writing and reading tasks: The retention effect on incidental second language vocabulary acquisition. *Journal of Education and Educational Research*, 13(1), 166-167. https://doi.org/10.54097/x8nh6e24.
11. Liu, S., & Reynolds, B. L. (2022). Empirical support for the involvement load hypothesis (ILH): a systematic review. *Behavioral Sciences*, 12(10), 354.

Nhận thức của người học về sử dụng Grok AI để lập dàn ý cho bài viết nghị luận

Nguyễn Hoàng Khâm*, Tô Vũ Lê Ngân, Lê Trần Minh Anh
Trung tâm Ngoại ngữ, Trường Đại học Nguyễn Tất Thành, Thành phố Hồ Chí Minh, Việt Nam
*nhkham@ntt.edu.vn

Tóm tắt Nghiên cứu này khảo sát nhận thức của người học tiếng Anh (EFL) ở trình độ B1 về việc sử dụng Grok AI, một chatbot dựa trên mô hình ngôn ngữ lớn (LLM), như một công cụ nhận thức phục vụ cho việc lập dàn ý bài văn nghị luận. Áp dụng thiết kế khảo sát cắt ngang sau can thiệp, nghiên cứu có sự tham gia của 128 học viên. Người tham gia đã sử dụng các câu lệnh được xây dựng dựa trên mô hình Toulmin và Giả thuyết Tải Tham gia (ILH) để tạo và đánh giá nội dung do AI tạo ra dưới sự hướng dẫn của giáo viên. Nhận thức của người học về tính hữu ích, tính dễ sử dụng và các yếu tố ILH được đánh giá thông qua một bảng câu hỏi thang đo Likert gồm 19 mục. Phân tích thống kê cho thấy người học có nhận thức rất tích cực trên tất cả các khía cạnh, trong đó việc lập kế hoạch cho minh chứng (data) được đánh giá cao nhất (*Mean* = 4.14), tiếp theo là lập kế hoạch cho luận điểm (claim) (*Mean* = 4.11) và lập kế hoạch cho cơ sở kết nối (warrant) (*Mean* = 4.08). Nghiên cứu này cung cấp những góc nhìn sơ bộ về việc ứng dụng mô hình nghị luận Toulmin vào việc sử dụng Grok AI, tập trung vào nhận thức của người học trong giai đoạn lập dàn ý trước khi viết. Dựa trên các kết quả này, một số cân nhắc sư phạm ban đầu về việc tích hợp AI vào giảng dạy viết EFL cũng được đưa ra thảo luận.

Từ khóa Tiếng Anh như một ngoại ngữ (EFL); nghị luận; Giả thuyết Tải tham gia (ILH); Grok AI.