WHIOCE PUBLISHING PTE. LTD.
PROVIDING
FIRST-CLASS SCIENTIFIC INFORMATION
FOR TOP SCHOLARS

2025 Volume 3, Issue 10
Educational Innovation Research
ISSN: 3029-1844(Print) ISSN: 3029-1852(Online)

The Effects of AI Assistance Granularity on Writers' Cognitive Load: An Empirical Study in Human-AI Co-Writing

Xinzhi Tao[1], **Jiaxin Li**[2]
[1]Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, China
[2]Northwest A&F University, Xianyang 712100, Shaanxi, China

Abstract: As more and more writers use large language models (LLMs) in their work, they face a big problem: how to enjoy the time savings that the technology brings without losing their own deep thinking and creative input. This is the key to making human-machine synergy truly effective. This paper methodically presents "AI-assisted granularity" as a key design element. By combining this idea with the cognitive load theory (CLT), we look at how it affects writers' intrinsic, extraneous, and germane cognitive loads in different ways. We modified the Leppink scale to create an AI-driven cognitive load assessment instrument for writing. We had 21 college students do argumentative writing tasks with AI help in three different ways: sentence-level, paragraph-level, and full-text structural guidance. The results show that full-text structural prompts greatly lower both intrinsic and extrinsic cognitive load, but they also make people less interested. On the other hand, prompts at the paragraph level are the best at increasing generative cognitive load, which shows a "moderate guidance" effect. On the other hand, sentence-level prompts make the brain work harder while giving the writer more creative freedom. At different levels of AI assistance, this study shows the ongoing conflict between "cognitive offloading" and "preserving agency." It suggests that writing tools in the future should have adjustable granularity mechanisms to make the best experiences for people and machines working together. These results offer a novel theoretical framework and design insights for AI writing research through the lens of cognitive load.

Keywords: AI-assisted writing, generative AI, human-computer interaction

Online publication: November 26, 2025

1. Introduction

Generative AI adds a new dimension of collaboration to writing, transforming it from traditional individual expression into a co-creation process between humans and machines. Although the advantages of AI are obvious - enhancing productivity and reducing psychological burden, its impact goes far beyond "assistance" itself.Large language models (LLMs), such as ChatGPT, have demonstrated outstanding performance as active "co-authors" in numerous natural language processing (NLP) tasks, including text generation, classification, summarization, language translation, and question answering[1]. This human-machine co-creation, while significantly enhancing writing efficiency, Marzuki's 2023 study[2] found that AI writing tools can help English learners write more effectively and improve their English proficiency. However, this also raises a core paradox: AI-assisted design, designed to promote productivity, may unintentionally undermine the core creative experience of authors.

AI-assisted systems can reduce both the intrinsic cognitive load resulting from task difficulty and the extrinsic cognitive load associated with using interfaces or tools, significantly by making it easier to find and combine information. This increase in efficiency, on the other hand, may come at a significant cost: it may unnecessarily suppress the generative cognitive load (related to internalizing knowledge, building mental models, and deep thinking), which is essential. When artificial intelligence gets too deeply involved, the writing process may degenerate from the "human-machine collaboration" of knowledge construction to a simple "task outsourcing", thereby sacrificing the author's process of learning, reflecting, and knowledge construction through writing.

The field of Human-Computer Interaction (HCI) has begun addressing this issue. Research demonstrates that while high-intensity AI "scaffolding" enhances output quality, it significantly reduces authors' sense of ownership and satisfaction with the process[3]. Previous studies treated AI intervention as a monolithic concept, overlooking how its effectiveness is influenced by specific implementation details. The granularity of AI-assisted design (i.e., the range of suggestions generated by AI: a single word, a complete sentence, or an entire paragraph) is a key but under-explored design dimension, which may create distinct cognitive and emotional experiences in balancing efficiency improvement with maintaining deep user engagement.

The research results show that different levels of artificial intelligence assistance (for example, sentence-level and paragraph-level suggestions) involve a clear trade-off between efficiency and user agency[4]. This trade-off stems from changes in cognitive load. However, the relationship between AI assistance granularity and users' cognitive load during usage remains unexplored. Understanding how different granularities affect the three components of cognitive load will enable us to provide optimal assistance strategies for diverse writing tasks and user goals—such as quickly replying to emails or drafting academic papers— thereby making informed choices between efficiency and deep thinking.

This paper aims to fill this gap. We propose that the granularity of AI assistance is a key mechanism moderating the complex trade-offs between authors' cognitive load, creative experience, and psychological engagement. Specifically, we hypothesize that higher-level assistance may maximize cognitive load reduction while simultaneously undermining user agency and flow experiences; conversely, lower-level assistance may better preserve these core experiences but offer relatively limited efficiency gains. Thus, understanding granularity effects is crucial for designing intelligent systems adaptable to diverse writing objectives.

2. Related work

2.1. The impacts of granularity on ai writing tools

The proliferation of AI-driven writing assistants has introduced a key design variable: the granularity of intervention. This refers to the scale and scope of feedback provided or content generated by AI systems, as well as the level of text interaction between AI tools and users. This factor not only affects the writing process but also the user experience, directly determining the role of AI in the collaborative relationship and profoundly influencing the user's cognitive experience and final output.

In recent years, the number of AI writing assistants has increased rapidly, with different designs in terms of the granularity of suggestions. Early intelligent writing tools were usually limited to automatic completion and error correction at the word or phrase level , while new assistants based on large language models (for example, ChatGPT, SudoWrite) can generate coherent multi-sentence or paragraphs-level text, thereby enabling larger-scale writing intervention. Tools like Wordtune mainly offer multiple revised versions of existing sentences, reflecting a more constrained sentence-level granularity (their core function is to provide rewriting options for the author's original phrase/sentence by adjusting the sentence structure or replacing synonyms)[5]. In contrast, generative models like the GPT series can draft or completely rewrite large amounts of text based on user prompts. Existing research has revealed how different intervention granularities affect the writing experience: sentence-level suggestions are often proven to be overly fine-grained, forcing authors to repeatedly switch between their own ideas and AI suggestions-disrupting the flow of thought and increasing cognitive load[9]. On the contrary, contributions of artificial intelligence at the paragraph level or higher can provide structural support during the writing process, which may reduce the cognitive burden in the early stage of creation and enhance attention. However, this advanced intervention is not without drawbacks. Although research shows that paragraph-level suggestions can significantly improve writing quality and efficiency, they can also reduce users' sense of control, ownership, and overall satisfaction. This phenomenon is known as the "AI ghostwriting effect", which highlights users' reluctance to claim that content mainly shaped by AI is their own.

In conclusion, granularity is a key factor influencing user experience. Our work introduces the concept of granularity to determine the optimal balance point for human-machine collaboration, achieve a balance among cognitive loads, and study how granularity systematically affects users' cognition and experience, thereby providing a new dimension for AI collaborative writing.

2.2. Cognitive and experiential factors in AI co-writing

John Sweller first came up with the Cognitive Load Theory (CLT) in the late 1980s [6]. Its goal was to improve instructional design and learning outcomes by matching the brain's cognitive abilities. Cognitive load theory categorizes cognitive load into intrinsic cognitive load, extraneous cognitive load, and germane/generative cognitive load. The main point of this theory is that "working memory" in people is very limited. It can only handle about four to seven independent pieces of information at a time. Cognitive overload happens when the amount of information that a task needs to process is too much for working memory. This makes it harder to learn and do tasks. Writing, as a typical complex cognitive activity, places extremely high demands on working memory. Therefore, CLT provides a powerful theoretical guidance for the analysis and design of AI writing tools. Sweller (1994) [7] noted that when the load exceeds the capacity, it hinders the fluency of writing. Therefore, writers (as well as tool designers) tend to improve efficiency by reducing the need for working memory. Nguyen et al. (2024) discovered that proficient writers synchronize various writing processes to prevent cognitive overload. Advanced AI can alleviate cognitive demands by automating routine tasks, enabling users to concentrate on higher-order structure and argumentation.

In recent years, the wide application of large language models (LLMS) in human-computer collaborative writing and digital online learning has highlighted the importance of cognitive load issues. CLT has always stressed that human working memory is limited and that we should reduce external loads and optimize related loads instead of just compressing internal loads. Studies show that AI systems can help students manage their cognitive load by giving them personalized feedback and step-by-step guidance, which can improve their learning outcomes PMC.

A comparative study of different writing styles (independent, human-led, and AI-led) found that compared with independent writing, AI-led and human-led styles significantly reduced cognitive load, especially in argumentative essays. In the context of artificial intelligence writing assistance scenarios, similar principles are still in the exploratory stage. The scaffolding method in human-machine collaboration is also closely related to cognitive load.Executed field experiments that compared three scenarios: the absence of AI, low granularity (e.g., sentence suggestions), and high granularity (e.g., paragraph suggestions). They discovered that high-granularity artificial intelligence suggestions markedly enhanced the quality and efficiency of writing, despite a minor reduction in text ownership and satisfaction; however, the cognitive burden did not experience a significant increase. It is worth noting that its impact on deep engagement and mental activities is beginning to emerge. The latest electroencephalogram (EEG) research from the MIT Media Lab shows that in three writing tasks, complete reliance on artificial intelligence assistance (such as ChatGPT) significantly reduced participants' brain connectivity (including alpha and beta waves), indicating a decrease in cognitive engagement - the so-called "cognitive debt" effect. Research in educational psychology suggests that AI design should ensure ongoing active cognitive engagement while minimizing cognitive overload. This can stop learning from turning into "knowledge task outsourcing" and instead encourage deep understanding through "germane load."

During the cognitive load scale measurement phase, Leppink developed the Cognitive Load Scale in 2013 [8]. This tool measures different types of cognitive load, and its robust psychometric properties have been validated in numerous studies. Developed and validated a Multidimensional Cognitive Load Scale for Virtual Learning Environments (VLE) (MCLSVE) in 2020, expanding and refining Leppink et al.'s (2013) Cognitive Load Scale (CLS). Shan Zhang et al. 2020) aimed to translate and validate the Chinese version of the Multidimensional Cognitive Load Tool (MDT-CL) for measuring different types of cognitive load, specifically to assess the cognitive load of Chinese ICU nurses [9].

The main goal of this paper is to look into how different levels of AI assistance affect cognitive engagement while writing and make the user experience better by lowering cognitive load. The Leppink scale was designed precisely to address such questions, moving beyond simple inquiries about "task difficulty" to delve into the sources of difficulty. Building upon the Leppink scale, our research created a novel AI-assisted writing cognitive load assessment scale.

The aim of design should be to augment, rather than supplant, human cognitive abilities should handle mechanical tasks, leaving higher-level structural and argumentative tasks—i.e., germane load—to humans, thereby enhancing writing ability without compromising skill development. Thus, the core objective of this study is not merely to measure the level of "total cognitive load," but to meticulously investigate how AI assistance at different granularity redistributes writers' limited cognitive resources from wasteful extrinsic load to productive relevant load.

3. Methods

We performed a field experiment to investigate the utilization of AI-driven collaborative writing tools and the impact of varying degrees of assistance from large language models (LLMs) on the co-authoring process. A screened group of 21 participants did argumentative writing tasks with different levels of AI support. Our personalized tool systematically gathered data by logging user inputs and utilizing various evaluation metrics.

3.1. Design of the experiment

Researchers gave participants an essay prompt from China's National College Entrance Examination (Gaokao) argumentative writing question bank for the experiment. We held a series of real-time virtual writing sessions through Tencent Meeting to create a controlled setting for collecting data. Participants (N = 1-3 per session) were directed through a hyperlink to a specially designed writing tool to finish their work (see **Figure 1**). A coordinator kept an eye on each session. To make sure the data was correct, we put in place key controls: (i) participants were kept apart in separate breakout rooms to stop them from interfering with each other; and (ii) mandatory camera activation and screen sharing verified that each task was done independently. This supervised, controlled process was meant to get accurate information about how participants write.

3.2. AI writing tool

For this experiment, we developed a customized collaborative AI writing tool. The frontend utilizes the Tiptap rich text editor framework integrated into a React application. GPT-4o is incorporated via the backend server into OpenAI's API endpoint, embedded within the workflow for text completion. Web interaction keys are as follows: - Tab: Request AI suggestions - Enter: Select and insert suggestions - Esc: Close suggestion panel - Up/Down arrows: Browse suggestion list The primary workflow (as shown in **Figure 2**) involves writing text, pressing Tab to seek AI suggestions, selecting suggestions to insert into the article, and then modifying or continuing the text.

3.3. Participants recruitment

We recruited 21 participants for the experiment through an online platform. We exclusively considered current undergraduate or graduate students residing in China, fluent in Chinese, aged 18-28, possessing basic AI usage experience, and demonstrating a certain level of cultural writing proficiency. Each task session was estimated to take 20 minutes to complete. Following the writing task, we assessed whether the writing quality met standards using AI-based writing scoring, with scores above 60 considered passable and usable data.

3.4. Data collaboration and measurements

We collect participants' responses to the Leppink Cognitive Load Scale adapted for AI-assisted writing experiments. Through questionnaires, we gather participants' actual experiences with AI-assisted writing at different granularity levels. The questionnaire assesses users' perceptions of intrinsic cognitive load, extrinsic cognitive load, and related cognitive load. All metrics are quantified using a low-to-high rating scale.

The session concluded with a 5-minute semi-structured interview. During the interview, experimenters asked questions to understand participants' observed interactions during the task, how they used the tool, and their experience using the tool in each scenario.

4. Empirical analysis

After collecting data from N = 21 participants, we proceeded with data analysis. We conducted a series of quantitative analyses using several statistical models discussed below.

Table 1. Descriptive statistics for cognitive load measures across experimental groups

Variable	Sentence Prompt (n = 21)	Paragraph Prompt (n = 21)	Text StructurePrompt (n = 21)
ICL_mean	4.44 ± 1.68	2.98 ± 1.19	2.75 ± 1.20
ECL_mean	4.86 ± 1.46	2.79 ± 1.29	2.10 ± 0.90
GCL_mean	3.38 ± 1.19	5.10 ± 1.32	4.87 ± 1.50

Note: Values are presented as mean ± standard deviation (SD). Group descriptions: Sentence Prompt = low-support condition; Paragraph Prompt = medium-support condition; Text Structure Prompt = high-support condition.

Table 2. ANOVA results for cognitive load by AI assistance granularity

Variable	Condition	F	P	Tukey HSD
Intrinsic Cognitive Load (ICL)	Full-Text	9.41	< 0.001	b
	Paragraph			b
	Sentence			a
Extraneous Cognitive Load (ECL)	Full-Text	28.23	< 0.001	c
	Paragraph			b
	Sentence			a
Germane Cognitive Load (GCL)	Full-Text	10.11	< 0.001	a
	Paragraph			a
	Sentence			b

Note: SD = Standard Deviation. For each variable, means in the same column that do not share a common superscript letter (a, b, c) are significantly different at $p < 0.05$ according to the Tukey HSD post-hoc test.

4.1. Intrinsic cognitive load (ICL)

The analysis indicated a substantial main effect of AI assistance granularity on ICL, $F(2, N-3)= 9.41$, $p < 0.001$. Post-hoc comparisons utilizing the Tukey HSD test revealed that the ICL in the Sentence condition (M = 4.44, SD = 1.68) was significantly elevated compared to both the Full-Text condition (M = 2.75, SD = 1.20) and the Paragraph condition (M = 2.98, SD = 1.19). There was no notable difference between the Full-Text and Paragraph conditions. This indicates that the provision of AI assistance at the sentence level enhances the perceived intrinsic difficulty of the task.

4.2. Extraneous cognitive load (ECL)

A substantial main effect of AI assistance granularity was identified on ECL, $F(2, N-3)= 28.23$, $p < 0.001$. Post-hoc tests indicated substantial differences among all three conditions. The ECL in the Sentence condition (M = 4.86, SD = 1.46) was the highest, and it was much higher than the ECL in the Paragraph condition (M = 2.79, SD = 1.29), which was also much higher than the ECL in the Full-Text condition (M = 2.10, SD = 0.90). This result shows a clear trend: as AI assistance becomes more detailed, it takes more mental effort to process the information.

4.3. Germane cognitive load (GCL)

There was a significant main effect of AI assistance granularity on GCL, $F(2, N-3)=$

10.11, $p < 0.001$. Unlike the patterns seen for ICL and ECL, post-hoc tests showed that the GCL in the Sentence condition (M = 3.38, SD = 1.19) was much lower than in the Full-Text condition (M = 4.87, SD = 1.50) and the Paragraph condition (M = 5.10, SD = 1.32). There was no significant difference between the Full-Text and Paragraph conditions. This indicates that more extensive forms of support (full-text and paragraph) enable learners to dedicate greater cognitive resources to schema construction and comprehension.

This study investigated the impact of artificial intelligence writing assistance at different granularity levels on authors' cognitive load. The results show that at the full-text granularity, the intrinsic cognitive load (ICL) and extrinsic cognitive load (ECL) reported by the participants were significantly reduced, while with the assistance of the sentence level, both types of load significantly increased. Similarly, as the support intensity increases, the external cognitive load significantly decreases, which is consistent with the view of the cognitive load theory that optimizing auxiliary design can reduce unnecessary cognitive needs. These findings support the research hypothesis that artificial intelligence assistance under the full-text structure minimizes the cognitive burden of writing.

However, in terms of relevant cognitive load, we observed an astonishing trend: paragraph level prompts generated the highest relevant load, followed by full-text structure prompts, while sentence-level prompts generated the lowest relevant load. This proves that when artificial intelligence provides moderate structural guidance (such as paragrapk-level prompts), authors can more easily focus their working memory on higher-order thinking and information integration, thereby increasing the related cognitive load. This "moderate guidance" effect is supported by research evidence on educational cognition (Kirschner et al., 2006)[10]. In contrast, sentence-level prompts are overly detailed, requiring authors to allocate most of their cognitive resources to local structures and coherence, which limits the space for in-depth processing. Although full-text structured prompts significantly reduce both intrinsic and extrinsic cognitive loads, highly structured information may lower the author's initiative and creative engagement - consistent with the self-determination theory that "excessive control reduces intrinsic motivation". Multiple studies have jointly confirmed that artificial intelligence support at an appropriate granularity can reduce cognitive load while stimulating deeper cognitive processing and participation, which is consistent with the findings of this paper.

The limitations of this study include: a controlled virtual meeting environment, participants completing only one argumentative essay, and a relatively small sample size (N = 21). Therefore, the universality of the results still awaits verification through larger-scale research and a more diverse range of writing tasks. Furthermore, indicators such as writing quality, creative output or author satisfaction were not evaluated in the current study. Future research can combine objective result evaluations and physiological indicators such as EEG to study the impact of different AI-assisted granularities on writing effectiveness and user experience in more realistic writing scenarios.

5. Limitations and future research

This study, based on a single controlled experiment, examined the effects of different levels of AI-assisted writing on the cognitive load of writers. The research results reveal the potential balance mechanism of granularity regulation in reducing both internal and external cognitive loads while maintaining generative cognitive loads. However, our research was influenced by several external variables.Firstly, the experimental task involved a one-time argumentative essay writing task, and the sample size (N = 21) was small, which limited the generality of the findings. Future research should validate these findings in larger-scale and more diverse writing tasks, especially in real educational and professional writing Settings.

6. Conclusion

Through field experiments, this study systematically investigated how artificial intelligence assistance at different granularity levels (sentence level, paragraph level, and full-text structure level) affects the three cognitive loads of authors. The results show that the full-text structure prompts significantly reduce both internal and external cognitive loads. Paragraph prompts perform best in reducing the cognitive load of association, while sentence prompts, although having a higher load, can reserve more space for autonomous conceptualization. This discovery provides empirical evidence for understanding the role of AI granularity in collaborative writing and offers important insights for designing human-AI collaborative writing tools - that is, flexibly adjusting the level of support according to the writing stage and user needs to balance efficiency, creativity and user engagement.

Disclosure statement

The author declares no conflict of interest.

References

[1] Wu T, He S, Liu J, et al., 2023, A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development. *IEEE/CAA Journal of Automatica Sinica*, 10(5): 1122–1136.

[2] Marzuki, Widiati U, Rusdin D, et al., 2023, The Impact of AI Writing Tools on the Content and Organization of Students' Writing: EFL Teachers' Perspective. *Cogent Education*, 10(2): 18.

[3] Dhillon PS, Molaei S, Li J, et al., 2024, Shaping human-AI collaboration: Varied scaffolding levels in co-writing with language models, Proc. [Conf. Abbrev.], [Page range].

[4] Fu L, et al., 2023, Comparing Sentence-Level Suggestions to Message-Level Suggestions in AI-Mediated Communication. *arXiv preprint arXiv:2304.05678*.

[5] Zhao X, 2023, Leveraging Artificial Intelligence (AI) Technology for English Writing: Introducing Wordtune as a Digital Writing Assistant for EFL Writers. *RELC Journal*, 54: 890–894.

[6] Sweller J, 1988, Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, 12(2): 257–285.

[7] Sweller J, 1994, Cognitive Load Theory, Learning Difficulty, and Instructional Design. *Learning & Instruction*, 4(4): 295-312.

[8] Leppink J, Paas F, Van der Vleuten CP, et al., 2013, Development of an Instrument for Measuring Different Types of Cognitive Load. *Behavior Research Methods*, 45(4): 1058-1072.

[9] Zhang S, Wu Y, Fu Z, et al., 2020, Psychometric Properties of the Chinese Version of the Instrument for Measuring Different Types of Cognitive Load (MDT-CL). *Journal of Nursing Management*, 28(2): 277-285.

[10] Kirschner PA, Sweller J, 2006, Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, 41(2).