SCIENTIFIC Publishing Limited

Journal of Qualitative Research in Education
https://ojs.ukscip.com/index.php/jqre

Article

Scaffolding the Story: A Mixed-Methods Investigation of AI-Generated Pre-Writing Support for EFL Continuation Writing

Zhengxian Li [1], Bao Chen [1], Ziying Ling [1], Rixi Wu [2] and Da Ling [1,*]

[1] School of Foreign Studies, South China Normal University, Guangzhou 510631, China
[2] School of Arts and Social Science, Monash University Malaysia Campus, Bandar Sunway 47500, Malaysia
[*] Correspondence: 20041054@m.scnu.edu.cn

Received: 8 December 2025; **Revised:** 27 December 2025; **Accepted:** 19 January 2025; **Published:** 28 February 2026

Abstract: This mixed-methods study investigates the efficacy and learner experience of using an AI-generated, multi-dimensional scaffold during the pre-writing stage of continuation writing tasks for Chinese English as a Foreign Language (EFL) learners. The quantitative phase employed a between-subjects design (N = 52 undergraduates), comparing an AI Scaffold Group with a No-Scaffold Control Group. Results indicated the scaffold group produced texts with significantly higher holistic quality, fluency, and lexical alignment, though no significant differences were found in grammatical accuracy. The qualitative phase, which is central to this study's purpose of achieving an in-depth understanding, involved stimulated recall interviews and scaffold use logs with a focal subsample. A reflexive thematic analysis of this rich qualitative data revealed that learners experienced the scaffold as a tool that: (1) reduced cognitive load by structuring the complex planning process, (2) deepened their comprehension and connection to the source narrative by activating a detailed situation model, and (3) increased confidence and self-efficacy by providing actionable linguistic and structural guidance. The integrated findings substantiate the role of AI-generated scaffolding as a potential instructional tool, demonstrating that its primary benefits are mediated through cognitive and affective pathways that facilitate a more engaged and meaningful interaction with the writing task. This study contributes to the fields of AI in education and L2 writing by delineating not only the outcomes but, more importantly, the lived experience and underlying learner-cognitive processes of AI-assisted planning, thereby highlighting the essential role of interpretive qualitative inquiry in evaluating educational technology.

Keywords: Continuation Writing; Pre-Writing Scaffolding; Artificial Intelligence; EFL Writing; Mixed-Methods Research

1. Introduction

Task-based language teaching (TBLT) frameworks commonly conceptualize the writing process as a sequence of three stages: pre-task, during-task, and post-task [1]. The pre-task stage, dedicated to strategic planning, is paramount. It is during this phase that writers formulate goals, generate ideas, and organize their thoughts to facilitate the actual composing process [1] (p. 62). Substantial evidence indicates that effective pre-task planning (PTP) helps learners manage limited attentional resources, activate relevant schemata and lexical knowledge, and ultimately yield more successful writing outcomes [2]. However, research into PTP has consistently documented a trade-off effect: while individual planning often enhances fluency and syntactic complexity, it frequently does so at the expense of grammatical accuracy, as learners struggle to simultaneously monitor meaning and form [3].

Collaborative planning, though beneficial for diversifying ideas, introduces its own complexities, including uneven participation and the potential reinforcement of interlanguage errors [4]. These persistent challenges underscore the necessity of exploring innovative, scalable approaches to planning support.

The investigation of PTP has been further limited by a narrow focus on discrete-skill tasks. In contrast to tasks that isolate reading or writing, the continuation task [5] integrates both modalities, requiring learners to comprehend a narrative text and then compose a coherent continuation. This integration is theoretically underpinned by the Interaction Hypothesis [6] and operationalized through the mechanism of alignment—whereby learners' linguistic representations converge through "dialogue" with the text at lexical, syntactic, and structural levels, driving L2 development [7,8]. The pedagogical promise of this task, however, is contingent upon learners' ability to deeply process the input text and strategically plan their continuation—a process that imposes significant cognitive and linguistic demands [9]. Learners often struggle with identifying narrative clues, generating creative yet coherent content, and selectively reusing sophisticated language from the input [10].

Instructional scaffolding, defined as temporary support that enables learners to accomplish tasks beyond their unassisted capacity [11], is a logical solution to these challenges. In continuation writing, well-designed pre-writing scaffolding can guide text analysis, idea organization, and linguistic noticing, thereby fostering both conceptual coherence and alignment [12]. Nonetheless, the provision of high-quality, individualized scaffolding by teachers is notoriously resource-intensive and difficult to standardize across diverse classrooms [13]. The recent advent of sophisticated Large Language Models (LLMs) like DeepSeek promises a transformative shift in AI algorithm development and utilisation. These models can generate adaptive, instant, and consistent scaffolding tailored to specific texts and learner profiles [14,15]. Preliminary research has demonstrated the benefits of structured planning support in continuation tasks [16], yet the empirical investigation of AI-generated scaffolds specifically targeting the pre-writing phase's unique cognitive-linguistic demands remains conspicuously sparse in existing literature. To address this gap, this study employs an experimental design to investigate the impact of a DeepSeek-generated, multi-dimensional scaffold on the continuation writing performance of Chinese EFL learners. The research specifically examines its effects on overall writing quality, fluency, accuracy, and the facilitation of alignment with the source text. And while quantitative measures can capture outcomes, a qualitative lens is essential to understand the cognitive and affective mechanisms through which such scaffolding operates. This study therefore employs an explanatory sequential mixed-methods design to not only measure the effect of the scaffold on writing performance but also to explore in depth the learner experiences, strategies, and perceptions that explain these effects.

2. Literature Review

2.1. The Continuation Writing Task and the Alignment Effects

Introduced by Wang (2012) [5], the continuation writing task requires learners to read an incomplete narrative and then write a logical and coherent conclusion. Its pedagogical power is largely derived from the alignment effect, a cognitive mechanism whereby interaction leads interlocutors to align their linguistic representations at various levels [7]. In the context of continuation writing, learners engage in a "dialogic" interaction with the source text, leading to alignment in lexicon, syntax, and narrative structure, which is hypothesized to facilitate implicit L2 acquisition [8]. This process aligns with broader second language acquisition theories that emphasize the role of input, interaction, and output in driving development [17].

A substantial body of research has confirmed the presence of alignment in continuation tasks. A meta-analysis of 19 experimental studies by Ren and Lyu (2021) [17] confirmed moderate alignment effects across various linguistic dimensions, affirming the task's utility across different educational stages. Furthermore, studies have identified several factors that modulate the strength of alignment, including text genre [18,19], linguistic complexity of the input [9], and specific task instructions [20].

2.2. Challenges in Pre-Task Planning for Continuation Writing

To effectively leverage the alignment effect in continuation tasks, careful attention must be paid to all three phases of task-based writing—pre-task, during-task, and post-task. Among these, the pre-task stage is particularly critical, as it is during planning that writers "establish goals for the writing, think up ideas related to these goals, and organize these to facilitate action" [1] (p. 62). Effective pre-task planning (PTP) helps learners manage competing

cognitive demands, activate relevant vocabulary and ideas, and ultimately produce more successful writing [2,21]. However, EFL learners face multiple cognitive and linguistic challenges during the planning phase of continuation tasks. At the conceptual level, many struggle to identify key narrative clues or foreshadowing in the original text, resulting in continuations that are logically inconsistent or thematically divergent [22]. Learners also frequently exhibit convergent thinking, producing highly similar storylines that lack creativity and diversity, thereby limiting the task's potential for language development [10].

Linguistically, learners often fail to notice or effectively reuse high-quality language from the input text. The linguistic complexity of input text has been shown to influence lexical complexity and fluency in continuation tasks [23]. Even when learners notice useful expressions, they may lack the proficiency to adapt them appropriately [24], undermining linguistic alignment with the original text [25]. This challenge in vocabulary reuse is further compounded by the inherent difficulty of mastering a large repertoire of lexical items, a core concern in vocabulary acquisition research [26].

Structural planning poses further difficulties. Many learners do not systematically plan their continuations using strategies such as mind-mapping or chronological sequencing [27]. As a result, their writing tends to be improvisational rather than strategically organized—a limitation that reflects broader challenges in macro-level thinking [21]. These planning challenges underscore the significant cognitive demands of continuation writing tasks, echoing findings that task complexity can profoundly affect learners' linguistic performance [28].

While collaborative planning can mitigate some of these issues [29], its effectiveness remains constrained by factors such as group dynamics, participants' proficiency levels, and the availability of qualified peers [30]. These limitations highlight the need for more individualized and accessible support, which instructional scaffolding can provide. Scaffolding refers to temporary and adaptive support that helps learners bridge the gap between their current abilities and the task demands [11]. In the context of continuation tasks, scaffolding during the pre-writing phase can assist learners in analyzing narrative clues, organizing plot structures, and noticing useful linguistic patterns from the input text, thereby enhancing both conceptual alignment and linguistic reuse [31]. The efficacy of such pre-writing scaffolding in addressing the specific demands of continuation tasks has been supported by recent mixed-methods research [12]. By offering structured guidance, scaffolding reduces cognitive overload and allows learners to focus on meaning and form simultaneously—key conditions for triggering alignment [8]. However, traditional scaffolding approaches often rely heavily on teacher expertise and intuition, making them labor-intensive, difficult to scale, and inconsistently implemented [26,32]. To address these practical challenges, AI-powered scaffolding presents a promising alternative. By leveraging natural language processing and adaptive learning technologies, AI can provide real-time, personalized planning support tailored to individual learners' needs, thereby making high-quality scaffolding more accessible and sustainable in continuation task settings.

2.3. AI-Assisted Scaffolding as a Potential Intervention

Large Language Models (LLMs) such as DeepSeek offer a promising means of supporting learners during the pre-writing stage. These models have demonstrated strong performance across various educational applications [14,33,34]. Empirical evidence underscores the critical role of pre-writing planning in specific tasks. In the context of continuation writing, the study by Xie and Zhu (2023) [15] demonstrated that providing structured planning modules (encompassing individual reading followed by individual or collaborative planning) directly addressed the challenges of generating linguistically complex and well-aligned narratives. Their findings revealed that such scaffolding, particularly individual planning, was effective in enhancing lexical and syntactic complexity and fostering linguistic alignment with the source text. LLMs can build upon this by generating even more sophisticated and tailored writing scaffolds that provide ideas, vocabulary, and structural guidance.

AI-generated support for integrated writing tasks can be conceptualized as comprising several key dimensions. In terms of content and ideation, AI scaffolding may assist learners in generating and organizing plot-consistent ideas [15]. Linguistically, it can provide lexical and syntactic suggestions aligned with the input text, thereby potentially enhancing both complexity and accuracy [16]. Structurally, it can offer guidance on narrative coherence and logical sequencing, which are critical for maintaining alignment with the source text [35]. However, the use of LLMs for scaffolding the pre-writing phase of continuation tasks remains underexplored. Most existing LLM writing tools focus on general writing support or post-writing feedback, rather than addressing the specific cognitive demands of continuation writing, which include deep comprehension, narrative consistency, and creative ideation [10]. Moreover, current tools often priori-

tize language-level assistance over guiding the underlying cognitive processes, which involve text interpretation, idea generation, and content organization and are essential for continuation tasks as they directly influence how learners comprehend, integrate with, and extend the source text [36]. Without scaffolding in these areas, learners may struggle to achieve meaningful alignment or produce creatively consistent and logically coherent continuations.

Thus, there is a clear need for research into how AI-generated scaffolds can be designed to address the cognitive, linguistic, and structural challenges of continuation task planning, while also fostering alignment and improving L2 writing outcomes.

3. Research Questions

To address the identified gaps, this study investigates the implementation and efficacy of DeepSeek-generated scaffolding in the pre-writing stage of a continuation writing task for Chinese EFL learners. It is guided by the following research questions:

(1) What are the core characteristics of the writing scaffolding generated by DeepSeek for the pre-writing stage of a continuation writing task, and how is it adapted to the cognitive and task demands of this phase?
(2) To what extent does the DeepSeek-generated scaffold enhance EFL university students' continuation writing performance, as measured by holistic scores, writing fluency, and writing accuracy?
(3) To what extent does the DeepSeek-generated scaffold promote alignment effects in the continuation writing process?

4. Methodology

4.1. Research Design

This study employed a between-subjects, quasi-experimental design complemented by a mixed-methods approach for data collection and analysis. The primary objective was to compare the continuation writing performance between an experimental group, which utilized an AI-generated pre-writing scaffold, and a control group, which engaged in traditional, unguided planning. The independent variable was the provision of the AI scaffold during the pre-writing phase, while the dependent variables encompassed holistic writing quality, textual fluency, grammatical accuracy, and the degree of linguistic alignment with the source text. Qualitative data were integrated to triangulate the quantitative findings and to provide deeper insight into the participants' cognitive processes and perceptual experiences with the scaffolding tool.

4.2. Participants

A cohort of 52 first-year undergraduate students was recruited from a public university in China. This study adopts a mixed-method approach. Quantitatively, as Lakens (2021) emphasized, a sufficient sample size is a prerequisite for valid statistical inference—our sample size thus supports the implementation of analyses (e.g., *t*-tests) to examine quantitative patterns [36]. Qualitatively, this design enables the selection of appropriate interviewees and the collection of in-depth, rich descriptive data, which complements the quantitative findings. The participant pool consisted of 26 females and 26 males, with a mean age of 18.9 years (SD = 0.7). All participants were native speakers of Mandarin Chinese and had received a minimum of ten years of formal English language instruction. To establish a baseline of proficiency, all individuals completed the Oxford Placement Test (OPT). Their scores placed them within the CEFR B1 to B2 range (M = 48.37, SD = 3.76). Participants were randomly assigned to either the AI Scaffold Group (n = 26) or the No-Scaffold Group (n = 26). An independent-samples t-test confirmed that no statistically significant difference existed between the two groups in their initial OPT scores ($t(50) = 0.48$, $p = 0.63$), thereby establishing group homogeneity in terms of general English proficiency prior to the intervention. For the in-depth qualitative inquiry, a purposeful sub-sample of 12 participants (6 from each group) was selected to maximize variation in post-test performance (high, medium, low) and gender. This sub-sample participated in stimulated recall interviews. This study has obtained ethical approval and the consent of the participants. The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of the School of Foreign Studies at South China Normal University. All participants were informed of the study's purpose, procedures, and confidentiality measures. Participation was voluntary, and informed consent was obtained from all subjects

involved in the study.

4.3. Materials

4.3.1. Post-Test Task

The post-test employed the continuation writing task from China's 2024 National College Entrance Examination (Gaokao) English Test Paper I. The text of this task presents a travel narrative full of twists and turns: it tells the story of "me" encountering taxi driver Gunter on a cold, wet, and unforgettable September evening. I was rushing to catch the last bus to Prague for a conference after my flight got delayed, and with Gunter's help, I barely made it to the bus station—only to fall into an awkward predicament: I had no cash on hand, my bank card failed to work, and the nearby cash machine was out of order.

Students need to sort out the emotional development trajectory of "me" (from anxious rush to relieved arrival, then to awkward helplessness), rationally design the plot based on the clues laid out in the previous text (such as Gunter's enthusiastic assistance, my bus-catching crisis and payment dilemma), and maintain the original language style through delicate psychological descriptions, vivid action depictions and natural interactive dialogues. The opening lines of the two continuation paragraphs are respectively, "I ran back to Gunter and told him the bad news." and "Four days later, when I was back in Vienna, I called Gunter as promised."

4.3.2. Intervention Materials

The AI Scaffold Group was provided with a pre-writing scaffold generated by DeepSeek-V3.1, developed in accordance with the general prompt outlined in Section 2.3. The API call was configured with the following parameters to ensure a balanced, informative, and deterministic output suitable for an educational scaffold: The temperature parameter was set to 0.7 to encourage coherent yet slightly varied phrasing. The maximum token output was limited to 500 to ensure a comprehensive yet concise scaffold. We employed a top_*p* value of 0.9, a frequency penalty of 0.5 to discourage excessive repetition, and a presence penalty of 0.3. A seed value of 42 was used to ensure reproducible generation across the research team's validation checks. The generated scaffold was reviewed by two EFL instructors for appropriateness and accuracy before administration, and no substantive changes were made to the AI-generated content. The prompt template used for generation is provided in **Appendix A**.

Following this generation process, the AI Scaffold Group was provided with the resulting pre-writing scaffold. This scaffold, totaling 382 words and calibrated to CEFR B1 level, comprised four modules: Situation Model Activation, which summarized the input text's core elements including characters, key events, setting, and unresolved plots; Linguistic Reuse Support, which extracted 5–7 reusable linguistic elements (vocabulary, phrases, and syntactic structures) each accompanied by an example of application in the continuation; Structure Guidance, which outlined a 3-step writing framework (opening → development → ending) with content and language prompts linked to the input text; and Cognitive Load Reduction, which identified common challenges for B1–B2 learners and provided corresponding solutions.

In contrast, the No-Scaffold Group received traditional pre-writing materials consisting solely of the input text and a blank outline, with no linguistic or structural guidance. This control group design was consistent with the protocol used in Peng et al.'s (2020) [22] study on continuation writing, ensuring comparability with existing research.

4.3.3. Assessment Tools

An integrated assessment system was developed to measure writing performance across multiple dimensions. The overall writing score was evaluated using a 25-point rubric that referenced Gaokao scoring standards and CEFR B1–B2 writing descriptors. This rubric mainly included four dimensions: Content Coherence, Structure Integrity, Language Appropriateness, and Task Completion. Two experienced senior high school English teachers, with 12 and 15 years of teaching experience respectively, conducted independent ratings to ensure objectivity.

Fluency was operationalized as the total number of words in the post-test essay, excluding punctuation and meaningless fillers, and was calculated using Range 32 software, following the measurement protocol established by Peng et al.'s (2020) [22]. Accuracy was indexed as the number of errors per 100 words, with five common error types coded: number agreement, article misuse, copula verb error, non-finite verb error, and tense error.

The measurement protocol follows the standardized framework established by Napoles, Nadejde, and Tetreault (2019) [37], which emphasizes the selection of core grammatical error categories and appropriate text length for robust evaluation. Specifically, the setting of five error types aligns with the core error categories recognized in their study, avoiding unreliable annotation caused by over-fragmentation of error types. Meanwhile, the 100-word text length is consistent with the recommended range of 80–120 words for GEC evaluation in their work, which has been empirically proven to effectively capture major error patterns without increasing annotation bias. Thus, our protocol is both scientifically grounded and practically feasible.

To measure alignment effect, we adopted Peng et al.'s (2020) [22] framework, focusing on the number of reusable linguistic elements from the input text in the continuation, including lexical alignment and phrase alignment, with overall alignment combining the former two metrics. AntConc 4.3.1 software [38] was used for vocabulary frequency analysis, while Notepad++ was employed for manual annotation of phrase alignment. This approach to linguistic annotation ensures a high degree of analytical precision, drawing on methodologies established in corpus-based linguistic research [39].

4.3.4. Supplementary Data Collection Tools

Supplementary tools were utilized to collect qualitative data and track scaffold usage. The AI Scaffold Group completed a paper-based scaffold use log immediately after writing, recording the modules they used, viewing duration, and marked linguistic elements. A stimulated recall interview guide with five semi-structured questions (e.g., "Which part of the scaffold did you find most helpful? Why?") was designed to explore participants' cognitive processes during the task. Additionally, a background questionnaire with 10 items was administered to gather information on English learning experience, proficiency self-assessment, and prior exposure to AI tools.

To gain an in-depth understanding of the participants' internal cognitive processes during the writing task, stimulated recall interviews were conducted with the 12 focal participants within 48 h of the writing task to capture fresh recollections. These 12 participants were purposively sampled to maximize variation in post-test writing performance (operationalized by holistic scores on the continuation task) and gender: within each group (AI Scaffold/No-Scaffold), 2 participants were selected from each performance tier (high: >75th percentile; medium: 25th–75th percentile; low: <25th percentile) based on post-test score distributions. This sampling strategy ensured comprehensive coverage of performance differences across groups, facilitating in-depth exploration of cognitive processes among diverse learners. The semi-structured interview guide was designed with open-ended, "how" and "why" questions to explore cognitive and affective processes deeply (e.g., "How did you decide which part of the scaffold to use first?" "Why did you choose to reuse those particular phrases?" "How did having the scaffold make you feel during the planning stage, compared to when you plan without help?"). Interviews were audio-recorded and transcribed verbatim, yielding over 200 pages of textual data for analysis.

All participants in the AI Scaffold Group also completed a structured log immediately after writing. This log prompted them to record which modules they used, for how long, and to annotate specific linguistic elements they marked for reuse. These logs provided tangible artifacts of engagement and were used as prompts during the stimulated recall interviews.

4.4. Procedure

The study was conducted over a 2-week period. In the first week, participants completed the background questionnaire (5 min) and were grouped based on their OPT scores. Both groups then received a 30-min training session on the requirements of the continuation writing task, with no linguistic or structural guidance provided to avoid confounding variables. The AI Scaffold Group used the DeepSeek-generated scaffold for 10 min of pre-writing preparation, followed by 45 min of writing, and completed the scaffold use log immediately after the writing task (5 min). The No-Scaffold Group spent 10 min of pre-writing preparation with the blank outline, followed by the same 45-min writing period. The 10-min duration allocated for scaffold engagement was intentionally designed based on two primary considerations. First, it aligns with established protocols in pre-writing intervention research, where brief, focused planning periods (e.g., 5–15 min) are commonly used to elicit strategic planning behavior without encroaching on the core writing time [1]. Second, this timeframe was calibrated to mirror real-world test conditions (such as the Gaokao), where planning time is often limited and integrated within a larger writing session.

In the second week, 12 focal participants (6 from each group) were selected for stimulated recall interviews,

which lasted 25–30 min per person and were audio-recorded for subsequent transcription and analysis.

4.5. Data Analysis

4.5.1. Reliability Analysis

Multiple reliability tests were conducted to ensure the validity of the assessment tools and coding procedures. For overall writing scores, the Intraclass Correlation Coefficient (ICC) was calculated to measure inter-rater reliability, while Cronbach's α was used for accuracy and alignment coding reliability, with a threshold of ≥0.90 set for acceptable consistency. For qualitative data, two researchers independently coded the interview transcripts using NVivo 14 [40], and the Kappa coefficient was calculated to verify coding consistency, with a threshold of ≥0.80.

4.5.2. Quantitative Analysis

Descriptive statistics, including mean (M) and standard deviation (SD), were calculated for all post-test variables. Inferential statistics were performed using SPSS 27.0: independent samples *t*-tests were used to compare group differences in post-test scores, and Cohen's d was calculated to measure effect size (with $d \geq 0.8$ defined as a large effect and $0.5 \leq d < 0.8$ as a medium effect). Pearson correlation analysis was conducted to explore the relationship between alignment and writing performance indicators.

4.5.3. Qualitative Analysis

The qualitative data analysis was guided by the principles of reflexive thematic analysis [41,42], an interpretive approach well-suited to exploring experiences, meanings, and processes. The analysis was primarily inductive, allowing themes to emerge from the data, but was also informed by the research questions. The process was iterative and involved six phases (**Table 1**):

Table 1. Steps and Instructions of Thematic Analysis.

Step	Instruction
Step 1 Familiarization	Repeated reading of all interview transcripts and scaffold logs.
Step 2 Generating Initial Codes	Systematic coding of interesting features across the entire dataset. Codes captured actions (e.g., "circling vocabulary"), perceptions (e.g., "feeling overwhelmed"), and reasoning (e.g., "chose phrase for versatility").
Step 3 Searching for Themes	Checking candidate themes against the coded extracts and the entire dataset to ensure they formed a coherent pattern and accurately reflected the data. This phase involved refining, splitting, or discarding themes.
Step 4 Reviewing Themes	Reviewing Themes: Checking candidate themes against the coded extracts and the entire dataset to ensure they formed a coherent pattern and accurately reflected the data. This phase involved refining, splitting, or discarding themes.
Step 5 Defining and Naming Themes	Developing a clear definition and name for each final theme that captures its essence.
Step 6 Producing the Report	Producing the Report: Selecting vivid, compelling extract examples to illustrate the analysis and weaving the analytic narrative together.

Trustworthiness and Rigor were ensured through several strategies: investigator triangulation (two researchers independently coded a subset of transcripts, achieving a Kappa coefficient of 0.85 before discussing and reaching consensus), data source triangulation (cross-checking themes emerging from interviews with evidence from scaffold logs), and maintaining an audit trail of analytical decisions.

4.5.4. Integration

Quantitative and qualitative results were integrated during the interpretation phase. The quantitative findings identified what effects occurred, while the qualitative themes provided a rich explanation for how and why those effects might have come about, offering a more complete understanding of the phenomenon.

5. Results

5.1. Reliability Analysis Results

Comprehensive reliability analyses were conducted to ensure the consistency and validity of all measurement instruments. The inter-rater reliability for holistic writing scores demonstrated excellent agreement, with an Intraclass Correlation Coefficient (ICC) of 0.96 (95% CI: 0.94–0.97, $p < 0.001$). Internal consistency for the coding

of accuracy and alignment measures was equally robust, yielding Cronbach's α coefficients of 0.95 (95% CI: 0.92–0.97) and 0.98 (95% CI: 0.96–0.99), respectively. For qualitative data analysis, inter-coder reliability reached a substantial level, as indicated by a Kappa coefficient of 0.85 (95% CI: 0.79–0.91). All reliability coefficients exceeded established thresholds, confirming the methodological rigor of the assessment procedures.

5.2. Group Homogeneity Analysis (OPT Scores)

Prior to examining intervention effects, group homogeneity in initial English proficiency was verified through analysis of OPT scores. As presented in **Table 2**, both groups demonstrated comparable proficiency levels at baseline. An independent-samples *t*-test confirmed no statistically significant difference between the AI Scaffold Group (M = 48.62, SD = 3.8) and the No-Scaffold Group (M = 48.21, SD = 3.72), with $t(50) = 0.48$, $p = 0.63$, Cohen's d = 0.13. This established equivalence ensures that any observed differences in post-test performance can be reasonably attributed to the experimental intervention.

Table 2. Descriptive Statistics and Homogeneity Test of OPT Scores.

Group	n	M (SD)	*t*-Value	df	*p*-Value	Cohen's d
AI Scaffold Group	26	48.62 (3.8)	0.48	50	0.63	0.13
No-Scaffold Group	26	48.12 (3.72)				

Note: M = Mean; SD = Standard Deviation; df = Degrees of Freedom; OPT = Oxford Placement Test.

5.3. Post-Test Group Differences

Descriptive Statistics and Inferential Results

The quantitative analysis revealed a distinct pattern of results across different dimensions of writing performance. As summarized in **Table 3**, participants in the AI Scaffold Group significantly outperformed their counterparts in the No-Scaffold Group on several key measures. Specifically, the scaffold group produced compositions of substantially higher overall quality, $t(50) = 3.53$, $p < 0.001$, Cohen's d = 0.98, and demonstrated markedly greater writing fluency, $t(50) = 3.51$, $p < 0.001$, Cohen's d = 0.97. In terms of linguistic alignment, the scaffold group showed significantly stronger lexical alignment, $t(50) = 3.69$, $p < 0.001$, Cohen's d = 1.02, and higher overall alignment with the source text, $t(50) = 2.58$, $p = 0.003$, Cohen's d = 0.72.

Table 3. Descriptive Statistics and Inter-Group Comparisons of Post-Test Indicators.

Indicators	AI Scaffold Group (n = 26)	No-Scaffold Group (n = 26)	*t*-Value	df	*p*-Value	Cohen's d
Overall Score	18.9 (14.46)	18.02 (2.49)	3.53	50	< 0.001	0.98
Fluency	208.27 (29.73)	195.31 (12.06)	3.51	50	< 0.001	0.97
Accuracy	3.38 (1.33)	3.65 (1.20)	0.77	50	0.45	0.21
Lexical alignment	3.42 (0.76)	2.65 (0.75)	3.69	50	< 0.001	1.02
Phrase alignment	2.35 (0.85)	2.31 (0.79)	0.17	50	0.87	0.05
Overall alignment	5.77 (0.99)	4.96 (1.25)	2.58	50	0.03	0.72

Note: Accuracy is quantified as the number of errors per 100 words. This metric exhibits a negative correlation with writing accuracy, where a lower value indicates a higher level of writing accuracy.

However, this advantage was not uniform across all measured dimensions. No statistically significant differences emerged between groups in writing accuracy, $t(50) = -0.89$, $p = 0.38$, Cohen's d = −0.16, or in phrase-level alignment, $t(50) = 0.17$, $p = 0.87$, Cohen's d = 0.05. This differential pattern suggests that the AI scaffolding intervention exerted selective effects, primarily enhancing content development and lexical appropriation rather than grammatical precision or the integration of multi-word units.

5.4. Correlation between Collaborative Effect and Writing Performance

Correlational analyses revealed significant relationships between alignment measures and key indicators of writing performance. As displayed in **Table 4**, overall alignment demonstrated strong positive correlations with both overall writing scores ($r = 0.68$, $p < 0.001$) and writing fluency ($r = 0.59$, $p < 0.001$). Conversely, a moderate negative correlation emerged between alignment and writing accuracy ($r = -0.53$, $p < 0.01$), suggesting that increased engagement with source text language may have occurred at the expense of grammatical monitoring.

Table 4. Correlation between Alignment and Writing Performance.

Indicators	Overall Score	Fluency	Accuracy
Overall alignment	$r = 0.68^{*}$ $p < 0.001$	$r = 0.59^{*}$ $p < 0.001$	$r = -0.53^{**}$ $p < 0.01$

Note: *** $p < 0.001$; ** $p < 0.01$.

5.5. Qualitative Results

5.5.1. Scaffold Use Log Analysis

Analysis of the scaffold use logs from the AI Scaffold Group (n = 26) provided quantifiable evidence of engagement preferences, which were further illuminated by participants' own explanatory notes. The "Linguistic Reuse Support" module was the most frequently accessed component, utilized by 85.7% of participants for an average of 3.2 min. This was not merely a passive viewing; logs indicated a strategic focus. As one participant annotated, "I circled 'deeply moved' and 'a sense of purpose' because they perfectly described Li Hua's feelings from the original text, and I knew I could use them again." This highlights a deliberate strategy of identifying and appropriating key affective vocabulary to maintain tonal alignment. Following this, the "Structure Guidance" module (used by 71.4%) and the "Cognitive Load Reduction" module (64.3%) were employed as organizational and problem-solving tools, respectively. In terms of concrete application, 78.6% of participants marked and subsequently reused at least four of the six provided linguistic elements. Notably, practical narrative phrases like "care for" and "share stories" saw the highest reuse rate. A participant's log comment clarifies this preference: "Phrases like 'care for' are versatile; I can use them to describe both actions and feelings in my continuation, which makes writing smoother." This suggests that learners prioritized reusable, functional language chunks that offered high utility for narrative development.

5.5.2. Stimulated Recall Interview Analysis

To ensure transparency in the qualitative analysis process, a systematic coding procedure was followed. Initially, two researchers independently open-coded a subset of four interview transcripts using NVivo 14, generating a preliminary codebook. Through iterative discussion and refinement, a final coding tree was established, organizing codes into sub-themes and overarching themes. An excerpt from the codebook is presented in **Table 5** to illustrate the derivation from raw data to themes. The full coding matrix is available as Supplementary Material. Themes were then richly illustrated with multiple verbatim excerpts from participants across the dataset, ensuring that each theme was grounded in and representative of the collected experiences.

Table 5. Excerpt from the Coding Tree Illustrating Theme Development.

Data Excerpt (Representative Quote)	Initial Code	Sub-Theme	Overarching Theme
"Normally my mind is a messy room when I start… This [scaffold] was like someone came and put up clear labels…"	"Metaphor of messy room vs. organized labels"	External organization of thinking	From Cognitive Overload to Managed Complexity
"After my first read, I had a general sense of the characters and setting, but it was all a bit vague… The summary in the scaffold focused my attention. It highlighted …"	"Sharpening vague comprehension"	Deepened text interpretation	Deepening the Dialogue
"The biggest help wasn't the words themselves, but the permission they gave me… It took away the fear of mistakes."	"Scaffold as permission-giver"	Reduced anxiety, increased risk-taking	Affective Scaffolding

Thematic analysis of the stimulated recall interviews revealed three core themes that elucidate the mechanisms behind the scaffold's impact for the AI Scaffold Group, while also highlighting contrasting experiences reported by the No-Scaffold Group.

Theme 1: From Cognitive Overload to Managed Complexity.

The Scaffold as an External Executive Function.

Participants from the control group characterized the unsupported planning phase as chaotic and overwhelming, using metaphors like "juggling too many balls at once" (P5, Control)—trying simultaneously to recall the story, invent a coherent plot, search for appropriate vocabulary, and monitor grammar. In contrast, participants from the scaffold group described how the AI scaffold transformed a typically demanding pre-writing phase into a more

structured process. For them, the AI tool served as an external organizer that imposed a manageable structure on this inherent complexity. It did not think for them, but rather structured their thinking for them.

> *"Normally my mind is a messy room when I start. I'm looking for the plot in one corner, vocabulary in another, and I can't find anything. This [scaffold] was like someone came and put up clear labels: 'Story Summary Here,' 'Useful Words Here,' 'Plot Steps Here.' I knew exactly where to look for what I needed. It didn't give me the answers, but it told me where to start digging." (P8, Scaffold)*

Interpretation: This theme highlights the scaffold's role in reducing extraneous cognitive load, a process critical for learners to allocate mental resources to higher-order tasks like idea generation and linguistic alignment.

Theme 2: Deepening the Dialogue: From Superficial Reading to Strategic Text Engagement.

For control group participants, interaction with the source text was predominantly described as a linear and limited process, often confined to an initial read-through to grasp the general plot. As one control participant noted, "I read it once to get the story, then I just had to start writing based on what I could remember. I didn't really go back to check details" (P2, Control). This approach sometimes led to uncertainty about character motivations or overlooked narrative clues, resulting in what another described as "a continuation that felt a bit disconnected from the original" (P5, Control). The scaffold group, in contrast, generally reported a more recursive and purposeful re-engagement with the text, mediated by the scaffold's prompts. The Situation Model Activation module served as a pivotal focusing lens, aiding in the construction of a coherent mental model of the story.

> *"After my first read, I had a general sense of the characters and setting, but it was all a bit vague. The summary in the scaffold focused my attention. It highlighted the underlying emotional tension and the main character's unresolved dilemma. It helped me see how I can develop my writing to respond to the original text, not just a separate story." (P7, Scaffold)*

Interpretation: This finding illustrates how the scaffold facilitated a deeper "dialogue" with the text, promoting the construction of a detailed situation model—a key cognitive process underpinning the alignment effect.

However, not all scaffold-group participants viewed this re-engagement as uniformly positive. A minority found the structured re-reading somewhat mechanical, with one participant commenting, "It felt a bit like checking boxes—I looked at the summary and the word list, but I'm not sure I really 'thought' more deeply about the story" (P9, Scaffold). Another expressed that while helpful, the scaffold's prompts "kind of told me what to focus on, so maybe I didn't come up with my own interpretation" (P12, Scaffold).

Theme 3: Affective Scaffolding: Building a Bridge from Anxiety to Agency.

In the control group, affective experiences were constrained. Participants frequently described feeling uncertain and apprehensive during planning. One stated, "I kept second-guessing whether my ideas fit the story or if I was even understanding it right" (P4, Control), while another explained, "I spent so much mental energy worrying about getting off track that I struggled to just start writing" (P2, Control). These accounts reflect a writing process more burdened by hesitation and self-doubt. In contrast, for the scaffold group, the scaffolding functioned as a psychological safety net, directly impacting learners' confidence and willingness to write.

> *"The biggest help wasn't the words themselves, but the permission they gave me. Without it, I'm scared to use a 'fancy' word in case I use it wrong. Seeing it listed as 'reusable' was like a teacher saying 'this is correct, you can use it.' It took away the fear of mistakes." (P11, Scaffold)*

Interpretation: This theme explains why learners in the scaffold group may have demonstrated greater fluency and lexical risk-taking. By reducing anxiety and building self-efficacy, the scaffold freed up cognitive resources and fostered a more agentic approach to writing.

Nevertheless, not all reactions in the Scaffold Group were uniformly positive. A few participants expressed neutral or reservations about the affective support. One noted, "It helped me feel less lost, but I still didn't feel truly confident—more like I was following a template" (P9, Scaffold). Another suggested a potential downside: "Knowing there was a 'right way' to use the vocabulary sometimes made me less creative. I stuck to the list even when I had

other ideas" (P12, Scaffold). These responses indicate that while the scaffold alleviated anxiety for most, it could also engender a sense of constraint or formulaic engagement for some.

5.6. Joint Display of Quantitative and Qualitative Findings

To illustrate the integration of quantitative and qualitative datasets, a joint display table (**Table 6**) is presented below. This table maps the key quantitative outcomes to the emergent qualitative themes, demonstrating how the qualitative insights explain and elaborate on the quantitative results.

Table 6. Joint Display of Integrated Quantitative and Qualitative Findings.

Quantitative Finding	Qualitative Theme	Interpretive Integration
Significantly higher holistic quality, fluency, and lexical alignment in AI Scaffold Group	Theme 1: From Cognitive Overload to Managed Complexity	Participants reported that the scaffold provided an external executive function, structuring the planning process. This reduction in extraneous cognitive load allowed them to allocate more attentional resources to idea generation and lexical selection, thereby enhancing fluency and overall text quality.
Strong lexical alignment but non-significant phrase-level alignment	Theme 2: Deepening the Dialogue: From Superficial Reading to Strategic Text Engagement	The scaffold facilitated focused re-engagement with the source text, especially through the Situation Model Activation module. This led to deliberate reuse of individual vocabulary items (lexical alignment) but did not sufficiently support the noticing or integration of multi-word phrases and syntactic patterns, explaining the lack of phrase-level alignment gains.
No significant improvement in grammatical accuracy	Theme 3: Affective Scaffolding: Building a Bridge from Anxiety to Agency	Learners reported increased confidence and willingness to take lexical and compositional risks, prioritizing meaning and fluency over grammatical form. This affective shift, while promoting fluency and alignment, may have diverted attention from grammatical monitoring, consistent with the trade-off hypothesis observed quantitatively.
Positive correlation between alignment and fluency/quality; negative correlation with accuracy	Themes 1 & 3 combined	The scaffold's dual role in reducing cognitive load (Theme 1) and lowering anxiety (Theme 3) enabled learners to engage more deeply with the source text and produce more fluent, aligned output. However, this focused engagement on meaning and lexical reuse likely reduced the attentional capacity available for grammatical accuracy, resulting in the observed trade-off.
Divergent & Nuanced Findings	Themes 1, 2 & 3 combined	Integrated analysis must account for divergence and differential experience. These minority responses suggest that while the highly structured AI scaffold provided a clear path for most learners, it may pose a potential constraint on the autonomous exploration and creative thinking of some. This offers important direction for future designs of more adaptive scaffolds that encourage diverse interpretations.

6. Discussion and Conclusion

6.1. Core Features of AI-Generated Pre-Writing Scaffolds

The findings of this study address the first research question by delineating three fundamental characteristics of the DeepSeek-generated pre-writing scaffold, which resonate with established sociocultural and cognitive learning frameworks. The first characteristic, proficiency adaptability, was manifested through the scaffold's construction using CEFR B1-level language and its alignment with the input text's complexity, ensuring accessibility without compromising the task's inherent cognitive demands. This finding corroborates Napoles et al.'s (2019) [37] emphasis on "input-output complexity matching".

The second characteristic, functional complementarity, was evidenced by the scaffold's four-module architecture targeting distinct pre-writing requirements—ranging from situation model activation to linguistic support, structural planning, and cognitive load reduction—thereby forming an integrated support system. This observation aligns with the "multi-dimensional scaffolding" paradigm in EFL writing pedagogy, which underscores the necessity of simultaneously addressing cognitive and linguistic barriers to optimize learning outcomes.

The third characteristic, text-linkage, was demonstrated through the scaffold's explicit connections to the input text via reusable linguistic elements and content prompts, thereby fostering interactive alignment between reading and writing processes. This substantiates Wang's (2016) interaction-alignment theory, which frames meaningful text-learner interaction as a critical mechanism for L2 development [43], and adds a critical empirical layer: AI scaffolds lower the threshold for alignment by making text-based linguistic resources more salient and accessible to learners with intermediate proficiency. The integration of these three features—proficiency adaptability, functional complementarity, and text-linkage—represents a significant advancement in AI-assisted writing scaffolding, providing a comprehensive framework that addresses the multifaceted challenges of continuation writing tasks.

However, the qualitative data reveal a nuanced fourth dimension crucial for understanding implementation: the tension between structured guidance and learner autonomy. While the integration of the three design features

explains the primary positive outcomes, the experiences of a minority of participants (e.g., P9, P12) highlight a potential constraint. Their feedback—feeling the process was somewhat mechanical ("like checking boxes") or that the scaffold limited creative interpretation—suggests that highly structured, directive scaffolding, while reducing anxiety and cognitive load for most, might inadvertently curtail exploratory thinking and personal ownership for some. This points to the dual-edged nature of structured AI support: it efficiently channels cognitive resources and builds confidence, but risks promoting a formulaic engagement if not designed or implemented with autonomy in mind [44].

Therefore, while quantitatively the scaffold improved holistic quality, fluency, and lexical alignment, qualitatively we understand this occurred because it fundamentally altered the planning experience for most learners by (1) managing intrinsic cognitive load through task segmentation (Theme 1), (2) focusing and deepening the text-interaction process (Theme 2), and (3) reducing anxiety and building self-efficacy (Theme 3). The lack of effect on accuracy aligns with the trade-off hypothesis and is further explained by participant reports: with reduced anxiety, they prioritized meaning and lexical risk-taking over grammatical form monitoring—a strategic shift enabled by increased confidence.

6.2. Effects of AI Scaffolds on Writing Performance

The second research question is addressed through the examination of group differences in post-test performance, which reveals a nuanced pattern of effects. The AI scaffold significantly enhanced overall writing quality, fluency, and lexical alignment with large effect sizes between experimental and control conditions—findings consistent with prior scaffolded writing research [37] and adding empirical weight to the value of AI-assisted writing interventions for continuation tasks.

The observed fluency enhancement can be attributed to the scaffold's capacity to reduce extraneous cognitive load through structured linguistic and organizational guidance, thereby enabling participants to allocate greater attentional resources to content development and text production. This interpretation aligns with Foster and Skehan's (1996) [45] limited attentional capacity model, which emphasizes the importance of cognitive resource management in writing performance. Furthermore, the observed trade-offs between different aspects of performance (e.g., fluency vs. accuracy) resonate with the predictions of the Trade-off Hypothesis and related models of task-based performance [45].

However, the non-significant findings for writing accuracy and phrase alignment warrant careful consideration. The absence of significant improvement in grammatical accuracy may reflect the trade-off hypothesis [45], whereby learners prioritize meaning over form when cognitive resources are constrained. Similarly, the lack of significant effects on phrase-level alignment suggests that while participants successfully reused individual lexical items, they encountered difficulties in incorporating multi-word sequences and syntactic patterns from the input text, possibly due to the greater processing demands associated with phrase-level integration [12]. While no significant group differences were found in accuracy, the negative correlation between alignment and accuracy suggests that increased alignment may trade off with grammatical accuracy in certain contexts.

In summary, while the quantitative results demonstrate the efficacy of the provided support, the integrated mixed-methods analysis allows for a more refined interpretation. The performance gains in quality, fluency, and lexical alignment likely stem from a combination of general benefits of structured planning (e.g., reduced cognitive load) and more specific affordances of the AI-generated scaffold (e.g., its adaptive, multi-faceted design that fostered deep text re-engagement and affective support). The pattern of results underscores that the scaffold functioned as a powerful enabling tool, but its effects were channeled through the inherent constraints and priorities of the L2 writing process, as evidenced by the persistent fluency-accuracy trade-off and the selective nature of alignment.

6.3. Mediating Role of Collaborative Effect (Alignment)

The third research question is elucidated through the significant correlations between alignment measures and writing performance indicators. The AI Scaffold Group's superior alignment performance, particularly in lexical alignment, indicates that the scaffold effectively facilitated deeper engagement with the input text. The positive correlation between alignment and overall writing quality substantiates the mediating role of alignment in the relationship between scaffolding and writing outcomes.

This finding reinforces Peng et al.'s (2020) [22] proposition that alignment constitutes a fundamental mecha-

nism linking input characteristics to writing development. The scaffold's explicit guidance for linguistic reuse appears to have lowered the threshold for alignment, enabling more effective text-learner interaction. Qualitative data further support this interpretation, with participants reporting active utilization of the scaffold's recommended elements, which contributed to both linguistic improvement and content coherence.

The alignment-fluency-quality pathway observed in this study provides empirical support for the theoretical mechanisms underlying continuation writing tasks and underscores the importance of fostering meaningful text-learner interaction in L2 writing pedagogy.

6.4. Theoretical and Practical Implications

6.4.1. Theoretical Implications

This study makes several significant theoretical contributions. First, it extends Wang's (2016) [43] interaction-alignment theory by demonstrating that AI-generated scaffolds can effectively enhance text-learner alignment in continuation writing contexts, thereby providing empirical validation for technology-mediated interactive learning in L2 acquisition. Second, it substantiates the application of cognitive load theory to AI-assisted writing contexts, illustrating how multi-dimensional scaffolding can optimize attentional resource allocation by reducing extraneous cognitive demands. These theoretical advances align with contemporary frameworks proposed by Sweller (2011) [44] and contribute to a more comprehensive understanding of the cognitive processes involved in AI-supported writing tasks.

6.4.2. Practical Implications

The findings offer several actionable implications for educational practice and tool development. For AI scaffold design, educators and developers should prioritize the three core features identified in this study—proficiency adaptability, functional complementarity, and text-linkage—with the four-module structure serving as a practical design template. In continuation writing instruction, teachers can strategically integrate AI scaffolds during pre-writing phases to support learners with limited task experience, for example, tasking students with using the AI tool to draft a thematic outline for their continuation based on the source text, then prompting them to modify the outline independently to foster original thinking and guard against over-reliance; this integration places particular emphasis on linguistic reuse and structural planning to enhance alignment between the source text and the continuation. For assessment purposes, the 25-point rubric and alignment measurement approach developed in this study can be adapted for standardized evaluation of continuation writing in diverse EFL contexts. However, potential risks—such as over-reliance on AI, reduced learner autonomy, or surface-level alignment without deeper knowledge acquisition—are acknowledged only briefly in existing discussions, highlighting the need for more targeted guidelines to mitigate such drawbacks in practical teaching.

6.5. Limitations and Future Directions

This study has several limitations that warrant consideration and suggest promising avenues for future research. The relatively small sample size (N = 52) constrains the statistical power and generalizability of the findings; compounded by the restriction to a single institution and a single task type (narrative continuation writing), these limitations further narrow the applicability of our conclusions. Thus, all claims should be carefully framed as context-specific to the current EFL setting and task design rather than broadly generalizable to diverse EFL learner populations or other AI-supported writing contexts.

The brief intervention duration precludes examination of long-term effects, highlighting the need for longitudinal investigations to explore the sustained impact of AI scaffolding on writing development. The 10-min duration for scaffold engagement, though designed to assess the tool's efficacy under pragmatic constraints and simulate its use in time-limited authentic assessment contexts, constitutes a key study limitation that directly undermined the depth of learners' cognitive engagement. This brief window forced participants to prioritize rapid completion of scaffold procedures over deliberate, reflective processing—particularly for metacognition-focused components such as the Cognitive Load Reduction module, which requires careful evaluation of AI-generated suggestions and alignment with personal writing goals. Learners had insufficient time to engage in iterative thinking (e.g., revisiting scaffold outputs to refine their writing plans) or articulate the rationale behind their choices, thus limiting the development of metacognitive awareness the scaffold was intended to foster. Consequently, the study may have

underestimated the scaffold's full potential. Future research should explore extended or self-paced scaffold interaction to clarify the relationship between engagement time, cognitive processing depth, and learning outcomes.

Additionally, the literacy levels of participants, including their reading comprehension and writing fluency in both L1 and L2, could be considered as a covariate or moderating factor in future studies. Individual differences in literacy may influence how learners interact with AI-generated scaffolds, particularly in tasks that integrate reading and writing, like continuation writing. Investigating whether scaffold effectiveness varies across learners with different literacy profiles could provide deeper insights into personalized scaffolding design and implementation.

Ecological validity represents another important consideration, as participants' awareness of being studied may have influenced their engagement with the scaffold. Future research should investigate scaffold effectiveness in authentic classroom settings to enhance ecological validity. Additionally, the exclusive focus on AI-generated scaffolding without comparison to traditional teacher-provided supports (e.g., vocabulary lists, writing frameworks) limits the precise delineation of AI's added value. Future studies should incorporate multiple comparison conditions to more accurately assess the relative benefits of AI-assisted scaffolding.

Promising research directions include investigating the influence of learners' AI literacy on scaffold utilization effectiveness, exploring the integration of AI scaffolds with other writing support tools (e.g., automated feedback systems), and examining differential effects across various learner profiles and proficiency levels. Such investigations would contribute to the development of more sophisticated, adaptive AI writing support systems that can effectively address the diverse needs of L2 writers.

Author Contributions

Conceptualization, methodology, writing—original draft preparation, and funding acquisition, Z.L. (Zhengxian Li) and B.C.; software, validation, and formal analysis, Z.L. (Ziying Ling); investigation, resources; data curation, R.W.; visualization, supervision, project administration, D.L. All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the Funding for Student Extracurricular Research Projects of South China Normal University, grant number 25WWKA01.

Institutional Review Board Statement

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of the School of Foreign Studies at South China Normal University.

Informed Consent Statement

Informed consent was obtained from all subjects involved in the study.

Data Availability Statement

To protect the participants' privacy, the data of this study cannot be shared openly.

Acknowledgments

The authors sincerely thank the participants for their cooperation and the reviewers for their constructive feedback, which has greatly improved the quality of this paper.

Conflicts of Interest

The authors declare no conflict of interest.

Appendix A. Prompt Template for AI Scaffold Generation

The following was the primary prompt template used to generate the multi-dimensional scaffold. Placeholders {input_text}, {cefr_level}, and {task_guidelines} were replaced with the actual source text (the 2024 Gaokao narra-

tive), "B1", and the standard continuation writing instructions, respectively.

You are an expert EFL writing tutor. Your task is to create a structured pre-writing scaffold to help a CEFR {cefr_level} learner plan a continuation for the following story.

STORY TO CONTINUE:

{input_text}

TASK: {task_guidelines}

Generate a comprehensive scaffold with the following FOUR sections. Use clear headings and bullet points. Use language appropriate for a {cefr_level} learner.

1. Situation Model Activation:

- *Summarize the KEY elements: main characters, their core traits/feelings, central conflict, setting, and any unresolved plot points.*
- *State the central theme or moral.*

2. Linguistic Reuse Support:

- *Extract 5–7 key lexical phrases, collocations, or sentence patterns FROM THE PROVIDED TEXT that would be highly useful for writing the continuation.*
- *For each item, provide a brief example of how it could be used in the continuation.*

3. Structure Guidance:

- *Provide a simple 3-part narrative framework (Beginning -> Middle -> End) tailored to this story's continuation.*
- *For each part, suggest 1–2 possible plot developments or questions the writer should consider.*

4. Cognitive Load Reduction Tips:

- *List 2–3 common challenges B1–B2 learners face in continuation tasks (e.g., "connecting the new plot to the old one," "reusing vocabulary naturally").*
- *For each challenge, offer one concrete strategy or tip.*

Ensure all guidance is directly linked to the provided story text.

References

1. Ellis, R.; Yuan, F. The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Stud. Second Lang. Acquis.* **2004**, *26*, 59–84. [CrossRef]
2. Hyland, K. *Second Language Writing*; Cambridge University Press: Cambridge, UK, 2010. [CrossRef]
3. Ong, J.; Zhang, L.J. Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *J. Second Lang. Writ.* **2010**, *19*, 218–233. [CrossRef]
4. Lee, J.; Burch, A.R. Collaborative planning in process: An ethnomethodological perspective. *TESOL Q.* **2017**, *51*, 536–575. [CrossRef]
5. Wang, C. The continuation task: An efficient way to facilitate L2 learning. *Foreign Lang. World* **2012**, *152*, 2–7. (in Chinese)
6. Long, M.H. The role of the linguistic environment in second language acquisition. In *Handbook of Second Language Acquisition*; Ritchie, W.C., Bhatia, T.K., Eds.; Academic Press: San Diego, CA, USA, 1996; pp. 426–468.
7. Pickering, M.J.; Garrod, S. Toward a mechanistic psychology of dialogue. *Behav. Brain Sci.* **2004**, *27*, 169–226. [CrossRef]
8. Wang, C.; Wang, M. Effect of alignment on L2 written production. *Appl. Linguist.* **2014**, *36*, 503–526.
9. Wang, C. The logic of promoting foreign language learning through continuation tasks. *Foreign Lang. World* **2021**, *207*, 2–7. (in Chinese)
10. Wood, D.; Bruner, J.S.; Ross, G. The role of tutoring in problem solving. *J. Child Psychol. Psychiatry* **1976**, *17*, 89–100. [CrossRef]
11. Jiang, Y.; Chen, J. The role of pre-writing scaffolding in L2 continuation writing: A mixed-methods study. *Lang. Teach. Res.* **2022**, *26*, 389–408.

12. Tai, H.-C.; Chen, C.-M.; Tsai, Y.-H.; et al. Is instructional scaffolding a better strategy for teaching writing to EFL learners? A functional MRI study in healthy young adults. *Brain Sci.* **2021**, *11*, 1378. [CrossRef]
13. Barrot, J.S. Using ChatGPT for second language writing: Pitfalls and potentials. *Assess. Writ.* **2023**, *57*, 100745. [CrossRef]
14. Lin, C.J.; Hwang, G.-J. Artificial intelligence-supported procedural scaffolding for promoting EFL learners' writing performance in flipped peer assessment activities. *Interact. Learn. Environ.* **2025**, *33*, 1–15. [CrossRef]
15. Xie, Y.; Zhu, D. Effects of participatory structure of pre-task planning on EFL learners' linguistic performance and alignment in the continuation writing task. *System* **2023**, *114*, 103008. [CrossRef]
16. Ellis, R. *Understanding Second Language Acquisition*, 4th ed.; Oxford University Press: Oxford, UK, 2019.
17. Ren, W.; Lyu, X. A meta-analysis of the alignment effect in the continuation writing task. *Foreign Lang. World* **2021**, *205*, 44–52. (in Chinese)
18. Zhang, X.; Zhang, L.J. Comparison of alignment effect in the continuation tasks of narration and argumentation. *Foreign Lang. World* **2017**, *177*, 90–96. (in Chinese)
19. Xin, S. Effects of continuation task condition on L2 grammatical structuring. *Mod. Foreign Lang.* **2017**, *40*, 507–517. (in Chinese)
20. Williams, J. *Teaching Writing in Second and Foreign Language Classrooms*; McGraw-Hill: New York, NY, USA, 2005.
21. Ellis, R. The effects of pre-task planning on second language writing: A systematic review of experimental studies. *Chin. J. Appl. Linguist.* **2021**, *44*, 131–165. [CrossRef]
22. Peng, J.; Wang, C.; Lu, X.F. Effect of the linguistic complexity of the input text on alignment, writing fluency, and writing accuracy in the continuation task. *Lang. Teach. Res.* **2020**, *24*, 364–381.
23. Kormos, J. The role of individual differences in L2 writing. *J. Second Lang. Writ.* **2012**, *21*, 390–403. [CrossRef]
24. Wang, C. From 'write-to-learn' to 'learn-by-continuation'. *Foreign Lang. Teach. Res.* **2017**, *49*, 547–556.
25. Nation, P.; Webb, S. *Researching and Teaching Vocabulary*; Pearson: New York, NY, USA, 2011.
26. Abrams, Z.I.; Byrd, D.R. The effects of pre-task planning on L2 writing: Mind-mapping and chronological sequencing in a 1st-year German class. *System* **2016**, *63*, 1–12. [CrossRef]
27. Kuiken, F.; Vedder, I. Task complexity and measures of linguistic performance in L2 writing. *Int. Rev. Appl. Linguist. Lang. Teach.* **2007**, *45*, 261–284.
28. Li, H.; Zhang, L.J.; Parr, J.M. Small-group student talk before individual writing in tertiary English writing classrooms in China: Nature and insights. *Front. Psychol.* **2020**, *11*, 570565. [CrossRef]
29. Kang, S.; Lee, J.-H. Are two heads always better than one? The effects of collaborative planning on L2 writing in relation to task complexity. *J. Second Lang. Writ.* **2019**, *45*, 67–72. [CrossRef]
30. Padmadewi, N.N.; Artini, L.P. Using scaffolding strategies in teaching writing for improving student literacy in primary school. In Proceedings of the 1st International Conference on Innovation in Education (ICoIE 2018), Padang, Indonesia, 28–29 September 2018; pp. 156–160. [CrossRef]
31. Wu, S.H.; Alrabah, S. Instructional scaffolding strategies to support the L2 writing of EFL college students in Kuwait. *Engl. Lang. Teach.* **2023**, *16*, 53. [CrossRef]
32. Su, Y.; Lin, Y.; Lai, C. Collaborating with ChatGPT in argumentative writing classrooms. *Assess. Writ.* **2023**, *57*, 100752. [CrossRef]
33. Zhao, X. Leveraging Artificial Intelligence (AI) Technology for English Writing: Introducing Wordtune as a Digital Writing Assistant for EFL Writers. *RELC J.* **2022**, *54*, 890–894. [CrossRef]
34. Sysoyev, P.; Filatov, E. Method of teaching students' foreign language creative writing based on evaluative feedback from artificial intelligence. *Perspect. Sci. Educ.* **2024**, 115–135. [CrossRef] (in Russian)
35. Qin, J. Relationship between learner creativity, task planning, and speaking performance in L2 narrative tasks. *Int. Rev. Appl. Linguist. Lang. Teach.* **2024**, *63*, 567–590. [CrossRef]
36. Lakens, D. Sample size justification. *PsyArXiv* preprint 2021. [CrossRef]
37. Napoles, C.; Nădejde, M.; Tetreault, J. Enabling robust grammatical error correction in new domains: Data sets, metrics, and analyses. *Trans. Assoc. Comput. Linguist.* **2019**, *7*, 551–566. [CrossRef]
38. Anthony, L. AntConc (Version 4.3.1) [Computer Software]; Waseda University: Tokyo, Japan, 2024. Available online: https://www.laurenceanthony.net/software/AntConc
39. Lu, X.; Wang, B.P.-Y. Towards a metaphor-annotated corpus of Mandarin Chinese. *Lang. Resour. Eval.* **2017**, *51*, 663–694. [CrossRef]
40. Lumivero. NVivo 15—The Most Trusted Qualitative Analysis Software (QDA) is Even Better. Available online: https://lumivero.com/products/nvivo/ (accessed on 10 October 2025).

41. Braun, V.; Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101. [CrossRef]
42. Braun, V.; Clarke, V. Reflecting on reflexive thematic analysis. *Qual. Res. Sport Exerc. Health* **2019**, *11*, 589–597. [CrossRef]
43. Wang, C.M. Promoting learning through "continuation": The continuation hypothesis of language acquisition. *Mod. Foreign Lang.* **2016**, *39*, 784–793, 873. (in Chinese)
44. Sweller, J. Cognitive load theory. In *Handbook of Writing Research*; Graham, S., Bus, A., Sterneberg, L., Eds.; Guilford Press: New York, NY, USA, 2011; pp. 396–407.
45. Foster, P.; Skehan, P. The Influence of Planning and Task Type on Second Language Performance. *Stud. Second Lang. Acquis.* **1996**, *18*, 299–323. [CrossRef]