https://doi.org/10.69639/arandu.v13i4.2706

Enhancing Academic Writing Organization Through AI-Supported Instructional Design: A backward design approach

Mejoramiento de la organización de la escritura académica a través del diseño instruccional apoyado en IA: Un enfoque de diseño inverso

Víctor Alejandro Briceño Ríos
vabricenor@ube.edu.ec
https://orcid.org/0009-0006-7591-4802
Universidad Bolivariana del Ecuador
Ecuador – Guayaquil

Rodrigo Antonio Guerrero Segura
raguerreros@ube.edu.ec
https://orcid.org/0000-0002-2698-5659
Universidad Bolivariana del Ecuador
Durán, Ecuador

Josué Reinaldo Bonilla Tenesaca
jrbonillat@ube.edu.ec
https://orcid.org/0000-0002-6748-2345
Universidad Bolivariana del Ecuador
Durán, Ecuador

Artículo recibido: 10 septiembre 2026- Aceptado para publicación:16 octubre 2026
Conflictos de intereses: Ninguno que declarar.

ABSTRACT

Academic writing organization is a critical competency in secondary education, yet many English as a Foreign Language (EFL) learners struggle with structural coherence and cohesion. This action research study evaluated the impact of an AI-supported instructional design, grounded in Backward Design and Communicative Language Teaching, to enhance paragraph writing organization among eighth-grade students ($N = 23$). Over a six-week intervention, participants engaged in structured drafting using ChatGPT as an automated formative scaffold to receive iterative feedback on paragraph structure, sentence cohesion, and logical coherence. A convergent mixed-methods approach was utilized, analyzing quantitative gains via pre- and post-intervention rubric assessments (t-test) alongside qualitative insights from an 18-item post-intervention Likert survey. Results revealed a statistically significant improvement in overall paragraph writing performance, with the class mean increasing from $M = 66.5$ ($SD = 16.4$) to $M = 87.0$ ($SD = 11.5$), $t(22) = 12.77, p < .001$, yielding an exceptionally large effect size ($d = 2.66$). The largest gains occurred in sentence cohesion and logical coherence ($d = 2.53$). Survey data ($M = 4.45, SD = 0.54$) indicated that learners perceived ChatGPT as an effective Socratic guide rather

than an automated writing surrogate. The study concludes that AI tools achieve maximum efficacy when intentionally embedded within a teacher-led, backward-designed curriculum.

Keywords: backward design, artificial intelligence, academic writing, formative feedback, secondary education

RESUMEN

La organización de la escritura académica es una competencia crítica en la educación secundaria; sin embargo, muchos estudiantes de inglés como lengua extranjera (EFL) enfrentan dificultades con la coherencia y la cohesión estructural. Este estudio de investigación-acción evaluó el impacto de un diseño instruccional apoyado en inteligencia artificial, fundamentado en el Diseño Inverso y la Enseñanza Comunicativa de Lenguas, para mejorar la organización de párrafos en estudiantes de octavo año de educación general básica ($N = 23$). Durante una intervención de seis semanas, los participantes realizaron actividades de redacción estructurada utilizando ChatGPT como un andamiaje formativo automatizado para recibir retroalimentación iterativa sobre la estructura del párrafo, la cohesión oracional y la coherencia lógica. Se empleó un enfoque mixto convergente, analizando las ganancias cuantitativas mediante evaluaciones con rúbrica pre y post-intervención (prueba t de muestras emparejadas) junto con el análisis de una encuesta Likert post-intervención de 18 ítems. Los resultados revelaron una mejora estadísticamente significativa en el desempeño general de la escritura de párrafos, aumentando el promedio de la clase de $M = 66.5$ ($SD = 16.4$) a $M = 87.0$ ($SD = 11.5$), $t(22) = 12.77, p < .001$, obteniendo un tamaño del efecto considerablemente alto ($d = 2.66$). Las mayores ganancias se registraron en las dimensiones de cohesión oracional y coherencia lógica ($d = 2.53$). Los datos de la encuesta ($M = 4.45, SD = 0.54$) indicaron que los estudiantes percibieron a ChatGPT como un guía socrático interactivo más que como un sustituto automatizado de redacción. Se concluye que las herramientas de IA alcanzan su máxima eficacia pedagógica cuando se integran deliberadamente dentro de un currículo guiado por el docente bajo el marco del diseño inverso.

Palabras clave: diseño inverso, inteligencia artificial, escritura académica, retroalimentación formativa, educación secundaria

Todo el contenido de la Revista Científica Internacional Arandu UTIC publicado en este sitio está disponible bajo licencia Creative Commons Atribution 4.0 International.

INTRODUCTION

Academic writing organization is a fundamental competency in middle school education. As students begin to develop different types of written texts, they must produce well-structured and coherent writing that demonstrates clear progression, coherence, and cohesion (Agustini et al, 2025). This stage is of paramount importance since it represents a key period for setting the foundation of writing organization for subsequent levels of education (Sarmiento et al., 2024). Furthermore, this pivotal ability not only helps students write more effectively but fosters critical thinking and communication skills across subjects (Graham et al., 2024). However, many learners struggle to organize their ideas effectively, often presenting unclear thoughts, limitations in linking devices, and weak paragraph structure (Siekmann et al., 2022). These difficulties affect writing in clarity and coherence. As a result, students compose texts that lack logical progression and clear communication of ideas. These challenges underscore the importance of implementing instructional approaches that explicitly target the development of writing organization skills and integrate innovative strategies, including AI-assisted tools, to enhance their abilities to structure their ideas more effectively.

In recent years, Artificial intelligence has been used in writing instruction as a means of supporting students' writing development (Kim et al., 2025). Recent studies have shown that the implementation of AI can provide immediate feedback, and support students during the writing process to develop text structure (Stitou & Fatmi, 2026). Also, the integration of AI in education for learning processes represents a significant advancement (Kasneci et al., 2023). Consequently, the use of AI-assisted tools plays a significant role in the students' skills as it functions as instructional support since it helps them revise their written texts in different ways to improve their writing quality in different components such as sentences, structures, cohesion and coherence (Ekizoğlu1 & Demir, 2025). According to Zhang, AI-generated content systems can significantly enhance student's writing skills when having structured feedback interventions (2025). Nevertheless, the integration of AI has been aimed at enhancing writing performance rather than specific components of writing (Li & Wilson, 2025). Previous studies also identified that scaffolding and structured guidance can positively enhance learners' writing production leading to well-organized and coherent texts (Baylon, 2023). Particularly, instructional approaches naming guided feedback, and structured questioning offers structured support helping students rearrange ideas, improve coherence, and structure paragraphs (Wei et al., 2023). Regarding the students' perceptions, the use of AI can transform their attitude into positive (Song & Song, 2023). From a curricular design perspective, these innovative strategies can be more effective when they are aligned within a backward design approach, which begins with clearly defined goals and aligns learning experiences and assessment accordingly ensuring that writing organization instruction are targeted in the learning process (Utami & Bram, 2023). Additionally,

research indicates that backward design supports the development of structured and coherent communication by organizing learning experiences into progressive and significant tasks (Zhang, 2026).

Despite the promising findings on AI-assisted tools in education, existing studies have primarily focused on how students can use AI to generally improve writing and provide automated feedback. However, research on the use of Chat GPT as a scaffolding tool in writing instruction to enhance writing organization in middle students remains limited. Furthermore, few studies have explored how AI-supported tools can be embedded within a backward design framework to support structured writing development. This gap highlights the need for further research that combines these elements to offer a more comprehensive approach by using this model.

This study aims to explore how AI-supported instructional design, based on a backward design approach, can enhance academic writing organization in eighth-grade learners. Specifically, it focuses on determining how the use of ChatGPT as an AI-supported instructional tool, within a backward design approach, can support students in structuring paragraphs, improve coherence, and strengthen their use of linking devices through the writing process, as well as to explore the students' perceptions and reaction of the use of AI during the intervention by using a pre-test and post-test and a survey with a Likert scale, and teacher field notes.

The significance of this study contributes to both education and research. The integration of AI tools like ChatGPT within an instructional design framework provides valuable insights regarding innovative strategies for improving writing organization. The findings of this research helped teachers implement an instructional model, effective teaching practices, and provide evidence on how AI enhances writing skills. Additionally, this study demonstrates the practical implications of using AI as a scaffolding tool to improve coherence, cohesion, structure and overall writing in students. Furthermore, it offered a foundation for future research on AI-supported instructional design approaches to enhance specific components of writing development across diverse educational contexts.

MATERIALS AND METHODS

Description of the intervention

The instructional intervention was implemented over six consecutive weeks and was structured according to the cyclical stages of Action Research: diagnosis, planning, action, observation and reflection. The instructional framework synthesized Wiggings McTighe's (Wiggins & McTighe, 2005) background design model with AI-scaffolded formative feedback delivered through ChatGPT. To ensure pedagogical implementation all AI prompts were designed in relation to the learning objectives of this research.

Stage 1: Diagnosis (Week 1). To start with the diagnosis and identify the students' gaps in academic writing, they had to complete a standardized paragraph-writing pretest following the

institution content. In this stage, there was no instructional support, teacher feedback, or AI assistance provided. The sole purpose of the assessment was to establish students' baseline performance in paragraph structure, sentence cohesion, and logical coherence.

Stage 2 Planning (Weeks 1-2). Based on the findings, a four-week instructional intervention was planned and implemented using Backward Design principles. Desire Learning objectives were identified at the beginning of each instructional cycle, assessment criteria were conducted through analytic writing rubric, and weekly writing tasks were structured. To constrain ChatGPT to a feedback role, teacher-designed templates were designed, ensuring the AI-assisted tool guided revision through targeted questioning rather than creating text for students.

Stage 3: Action (Weeks 2-5). The following intervention was implemented over four weekly sub-cycles.

- **Week 2 (Paragraph Structure):** This week was focused on the first part of the instruction which is the paragraph structure, containing supporting details, and concluding sentence to produce organized writing. Students submitted initial drafts to ChatGPT to analyze their paragraph models, receive formative feedback, and revise their own writing.
- **Week 3 (Sentence cohesion):** The focus of this week was on linking devices, reference words, and lexical variety. What ChatGPT provided was immediate feedback on how well these three elements were used in their paragraphs to evaluate and make future changes on their paragraphs.
- **Week 4 (Logical coherence):** This week had the objective of organizing ideas, making their sentences more coherent and easy to follow. Students used pre-structured AI-supported prompts to detect inconsistencies, eliminate abrupt transitions, and improve the overall flow of information or analysis.
- **Week 5 (integrated synthesis):** Students completed full expository paragraphs combining the organizational components. The use of ChatGPT was restrained to be used exclusively as an automated formative scaffold, asking reflective questions and identifying weaknesses in their paragraphs. This way students would know what they need to improve on.

Throughout this stage, teacher participation was essential to guide students, model writing strategies, monitor students' progress intervention, and assist each student with the use of ChatGPT as a scaffolding tool.

Stage 4: Observation (Weeks 1-6). In the implementation, observation was systematically applied to track student engagement, monitor their interaction with the AI-assisted tool (ChatGPT), and record pedagogical dynamics. Observation was conducted along with teacher field notes, and monitoring of student chat logs to ensure that AI served strictly as a scaffold rather than a tool to provide elaborate texts.

Stage 5: Reflection (Week 6). To finish the intervention, students had to complete an equivalent paragraph-writing posttest under parallel conditions set during the pretest.

Immediately afterward, participants also completed a perception survey using a Likert scale designed to observe their learning experiences using ChatGPT as an AI-supported instructional tool during the writing process.

Research design

This study adopted a convergent mixed-method approach, including quantitative and qualitative data to obtain a comprehensive understanding of the effect an AI-supported instructional intervention on students' academic writing organization. Mixed-methods research combines the strengths of quantitative and qualitative approaches, allowing researchers to examine both measurable learning outcomes and participants' perceptions of the educational experience (Creswell & Creswell, 2023).

Methodologically, the investigation utilized classroom-based Action Research to enhance an authentic teaching practice through a structured cycle of planning, action, observation, and reflection (Burns, 2010). To assess the impact of the research in students' writing organization, a quasi-experimental one-group pretest–posttest design was adopted. No control group was involved since all participants belonged to the same class.

The pedagogical intervention was designed based on the principles of Backward Design (Wiggins & McTighe, 2005), beginning with the identification of desired learning outcomes, next followed by the selection of assessment evidence and lastly the planning of instructional activities. The use of AI (ChatGPT) was incorporated as scaffold throughout the intervention, obtaining formative feedback, guiding students' revision processes, and encouraging metacognitive learning without generating complete written products. Total responsibility and accountability were taken by the teacher ensuring that the AI-supported tool functioned as educational support.

Research participants

The study was conducted at a private educational institution located in Guayaquil, Ecuador. The intervention took place during regular English as a Foreign Language (EFL) writing lessons over a six-week instructional period during the 2026 academic year. The target population consisted of eighth-grade students enrolled in the educative institution in Language Arts as the English subject. Since this Action Research seeks to enhance a specific educational context, a non-probability convenience sampling strategy was adopted. The final sample consisted of twenty-three students (N = 23) from a single intact classroom. All Participants shared the same instructional conditions, including the same teacher, syllabus, class hours, and English proficiency level, which minimized the risk of potential external variables.

All students' participation was voluntary across the different stages and phases of the research process. With no students' dropouts, the study achieved a 0% attrition rate (N:23) maintaining the same number of students until the end of the research. The relatively homogeneous characteristics of the sample ensured uniform implementation of the instructional

framework, thereby facilitating a rigorous comparison of student writing organization outcomes (pre-test and post-test).

Data collection procedures

Data collection spanned a six-week period following the implementation of the AI-supported instructional intervention. Quantitative data were gathered using a paragraph-writing pretest and a writing posttest rubric, whereas qualitative data were collected via a student perception survey administered at the conclusion of the study.

During **Week 1**, participants completed the writing pretest to analyze and establish their baseline level of writing organization. The pretest was administered during regular English lessons under standardized classroom conditions as the students followed their subject content. Participants individually wrote an expository paragraph of 120–150 words within 30 minutes without any assistance. Upon completion, each writing sample was assigned an anonymous identification code (S01–S23) to ensure participant confidentiality throughout the research process and according to the school policies.

Between **Weeks 2 and 5**, the four-week Background Design intervention was implemented as previously detailed. During this period, no formal testing occurred, the researcher-maintained field notes to monitor students' engagement, track attendance and verify the implementation of the instructional procedures.

During **Week 6**, participants completed the writing post-test under identical conditions as the pretest. An equivalent expository writing prompt was used to keep task difficult while mitigating testing and practice effects. Students completed the assessment independently with no human or AI assistance, ensuring that the resulting texts represented their individual writing performance after the intervention.

Right after, following the posttest session, participants complete the Student Perception Survey. The questionnaire comprised sixteen closed-ended items measured on a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). Students filled out the survey anonymously in approximately 10–15 minutes under teacher supervision.

Scoring Procedures

All baseline and post-intervention writing samples were analyzed and evaluated using the same analytic writing rubric to ensure consistency. The scoring instrument assessed three components of writing organization: paragraph structure, sentence cohesion, and logical coherence. Paragraph structured was weighed at 40 points, while coherence and cohesion at 30 points resulting in a maximum total score of 100 points. Based on the total score, the grading performance was classified into 4 levels: Advanced (100-90), Proficient (75-89), Developing (60-74), and beginning (1-59). Individual scores for each dimension and the total writing organization grade were recorded in a Microsoft Excel spreadsheet for a better analysis. The use of a single

scoring instrument for both assessments (pretest and posttest) ensured consistency and comparability.

Coding of Questionnaire Responses

Participant responses from the Perception Survey were numerically registered prior to analysis using a 5-point Likert scale. The values are the following for each response option: 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, and 5 = Strongly Agree. Each item in the survey was coded based on the corresponding dimensions for the dependent and independent variable: Paragraph Structure (Items 1–3), Sentence Cohesion (Items 4–6), Logical Coherence (Items 7–9), Usefulness and Interaction (Items 10–12), Scaffolded learning (Items 13-15) and Instructional design efficacy (Items 16–18).

Instruments

The data obtained in this research relied on three primary instruments to collect both quantitative and qualitative data: (a) a writing pretest, (b) a writing posttest, (c) a student perception using a five-point Likert scale, and (d) field notes during the process. The pretest and posttest writing assessment intervention generated quantitative data to evaluate gains in students' writing organization. Whereas the perception survey was used to capture qualitative data regarding students' reflections and experiences with the instruction design implemented through ChatGPT as a scaffolding tool, and field notes

Writing Pretest

The pretest served as diagnostic test to measure students' writing organization before the intervention. Under standardized classroom conditions, all participants completed an individual paragraph-writing task without requiring any type of assistance. Students task required to write a well-organized expository paragraph of approximately 120–150 words within a maximum time of 30 minutes in response to a writing prompt appropriate for their English proficiency level which was “Do social media platforms have a positive or negative impact on teenagers?”. The prompt required students to construct a single main idea supported by the core topic, supporting details and an appropriate concluding sentence, as well as the use of connectors. The procedures were identical for all participants. Students did the activity individually during a regular schedule, and no dictionaries, online resources, or AI-powered tools were allowed. The objective of this pretest was to establish the participants’ initial performance in the three dimensions previously mentioned which are: paragraph structure, sentence cohesion, and logical coherence.

Writing Posttest

The posttest was administered during the final week of the study following the AI-supported instructional activities. To mitigate testing effects, students participated in a different writing prompt that required the same writing genre, length, linguistic complexity, and organizational skills assessed in the pretest. As in the initial assessment, students composed an expository paragraph consisting of approximately 120–150 words within a 30-minute period

under the same classroom conditions answering the writing prompt “Does using artificial intelligence tools for schoolwork help or hind students’ learning?”.

Participants did not receive instructional support, teacher feedback, or AI assistance during the assessment to ensure that the resulting texts reflected students' independent writing performance. The administered procedures allowed direct comparison between pretest and posttest scores, making it possible to determine changes in writing organization following the intervention. Both assessments (pretest, and posttest) were evaluated under the same analytic rubric to ensure consistency throughout the research process.

Analytic Writing Rubric

Students’ written tasks were evaluated using analytic scoring rubric specifically designed to assess writing organization. Instead of assigning a score in a holistic way, the rubric measured the three organizational components established as the dependent variable of the study. The rubric was adapted and modified according to the need of the research objective from internationally recognized writing assessment frameworks, including East (2009), Halliday and Hasan's (1976) theory of cohesion, and recent research on coherence in academic writing (Riswanto, 2021). The implementation and adaptation ensured that each assessment criteria aligned directly with the instructional objectives established through the Backward Design framework.

The rubric consisted of three dimensions:

- **Paragraph Structure (40 points)**: Evaluated the quality of the topic sentence, supporting details, unity, and concluding sentence.
- **Sentence Cohesion (30 points)**: Assessed the effective application of cohesive devices, transition signals, lexical repetition, reference chains, and grammatical connections.
- **Logical Coherence (30 points)**: Examined the logical sequencing of ideas, clarity of progression, argumentative consistency, and overall paragraph readability.

The rubric showed a maximum composite score of 100 points. Based on the raw composite scores, the overall writing was divided into four performance tiers: Beginning (1.00-59.99), Developing (60.00-74.99), Proficient (75.00-89.99), and Advanced (90.00-100).

To establish content validity, the rubric criteria were aligned with the study's operational definitions and target learning outcomes. Prior to data collection, the rubric was reviewed by a specialist in English language teaching and academic writing assessment to ensure it was aligned to the study outcomes. The reviewer examined the clarity, relevance, and alignment of each descriptor with the research variables and provided feedback that was incorporated into the final version of the instrument.

To enhance scoring reliability, all writing samples were evaluated using identical scoring criteria and standardized procedures in the pretest and posttest.

Student Perception Survey

After intervention, students participated in a perception survey designed to examine their learning experiences with ChatGPT as an AI-supported instructional tool during the writing process. The survey comprised eighteen statements measured using a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). The survey was aligned with the conceptual framework of this research containing 5 dimensions. The instrument was employed across six distinct dimensions to analyze students' perceptions on writing organization, including paragraph structure (Items 1–3), sentence cohesion (Items 4–6), and logical coherence (Items 7–9). The other dimensions examined perceptions of the AI as an independent variable, including usefulness and interaction (Items 10–12), Scaffolded learning (Items 13-15) and Instructional design efficacy (Items 16–18) as detailed bellow.

Dependent Variable: Writing Organization

Paragraph structure

1. ChatGPT helped me understand how to format a single paragraph using a clear topic sentence.
2. The interactions with ChatGPT guided me to add relevant and detailed supporting sentences to my core idea.
3. Using the AI tool helped me write better concluding sentences that neatened the summary of my text.

Sentence cohesion

4. ChatGPT suggested varied transition words (e.g., *furthermore, in contrast, consequently*) that improved my text connectivity.
5. The feedback from the AI tool helped me replace repetitive or mechanical linkers like "and", "then", or "because".
6. Practicing with ChatGPT helped me make my pronoun references clearer, avoiding confusion between sentences.

Logical Coherence:

7. ChatGPT guided me to organize my thoughts in a highly logical, easy-to-follow sequence.
8. The AI scaffolding prevented me from making abrupt jumps or losing the main thread of my argument.
9. The tool helped me review my text so that the reader does not have to pause to infer my train of thought.

Dependent Variable: A

Usefulness and interaction

10. The formative feedback provided by ChatGPT was easy to understand and clear during the writing process.
11. ChatGPT acted as an effective academic guide, helping me improve my drafts rather than writing the text for me.

12. Receiving immediate recommendations from the AI tool made the instructional design more interactive and useful.

Scaffolded Learning

13. The guiding questions and prompts provided by ChatGPT helped me think about how to organize my ideas before writing.
14. Using ChatGPT encouraged me to revise and improve my own writing instead of simply accepting the first version.
15. ChatGPT acted as a learning guide that supported my writing process rather than completing the work for me.

Instructional design efficacy

16. The AI-supported instructional activities helped me achieve the learning objectives for paragraph writing.
17. The AI-supported instructional design provided clear guidance throughout each stage of the paragraph writing process.
18. The combination of teacher instruction and AI support made the paragraph writing activities more effective for my learning.

Each statement was directly formulated from the operationalized variables to ensure alignment between the research objectives, instructional intervention, and data collection instruments.

Teacher Field notes

The use of field notes was implemented to keep evidence of the students' progress and to record how the study was advancing.

Stage 1: Diagnosis (Week 1)

Pretest assessment: Students completed the pretest without teacher guidance. They did not have any device nor the assistance of AI.

Observation: Students showed some frustration when they had to use connectors to keep the paragraph flow.

Stage 2: Planning (Weeks 1-2)

Observation: The prompts design for this study is structured in a way it can provide feedback and guidance, rather than a text generator.

Stage 3: Action (Weeks 2-5)

Week 2 Paragraph Structure

Observation: Students started drafting their paragraphs to submit it to ChatGPT to evaluate the structure of their paragraphs. The AI gave them feedback on their structure to align their supporting details to the main idea.

Week 3 Sentence Cohesion

Observation: Students actively used ChatGPT to analyze how they can use linking devices to keep continuity in their paragraphs. Some students replaced changed the words “and”, “so”, with more academic connectors such as “Therefore”, “Besides”.

Week 4 Logical Coherence

Observation: Using pre-structured prompts, students queried ChatGPT to identify abrupt transitions and logical gaps. Learners were re-ordering sentences and removing redundant statements to improve paragraph flow.

Week 5 Integrated Synthesis

Observation: Students integrated the three organizational elements into an expository paragraph. The use of ChatGPT functioned as automated scaffold providing reflective analysis on weak areas. Teacher monitoring confirmed they were using to receive feedback and not copying texts.

Stage 5: Reflection (Week 6)

Observation: Students completed the posttest assessment without using ChatGPT. Learned demonstrated much more confidence and efficacy during the posttest. They also expressed a positive attitude towards the use of the tool by completing the survey.

All these observations were gathered during the 6 weeks as specified in stage 4 to record how they were working, but at the same time how they responded to the implementation of this study.

Data Analysis

Data from this study were processed quantitatively and qualitatively. Raw data obtained from the pretest-posttest assessments and the survey were coded in an Excel spreadsheet and then transferred to the SPSS program to analyze them quantitatively and generate descriptive and inferential statistics. Qualitative observations recorded in teacher field notes across the five action research stages were processed through thematic analysis to triangulate the empirical findings.

Descriptive statistics were computed to summarize the quantitative datasets and provide a global representation of baseline and post-intervention student performance. Central tendency and dispersion metrics including mean (M), standard deviation (SD), minimum, and maximum scores were calculated for overall composite results as well as for each specific rubric domain: Paragraph Structure, Sentence Cohesion, and Logical Coherence. For the student perception survey, itemized means, standard deviations, and overall agreement percentages were aggregated.

Inferential statistics included a paired-samples t-test to determine whether score differences between the pretest and posttest ($N = 23$) were statistically significant. All testing was conducted at a 95% confidence level with a standard significance threshold of $\alpha = .05$. Finally, Cohen’s d effect sizes were calculated to measure the practical impact of the intervention. Effect size values were interpreted using standard thresholds: $d = 0.20$ as a small effect, $d =$

0.50 as a medium effect, and $d \geq 0.80$ as a large effect. Observational field notes were categorized by themes to explain how students interacted with ChatGPT during each stage of the study.

Ethical Considerations

The study was conducted in accordance with the ethical principles governing educational research involving human participants. Before data collection, Formal authorization was granted by the head of the school to conduct the study. Parents or legal guardians provided informed consent for students to participate in the research, while students gave their assent after receiving a clear explanation of the study's objectives and procedures.

In this study, participation was entirely voluntary, guaranteeing the right to withdraw at any moment or stage during the intervention without any academic consequences. To maintain participants' privacy, all writing samples and survey responses were anonymized using identification codes, and no personal information was included.

Limitations

Various limitations should be considered regarding the findings of this study. First, the research conducted a non-probability convenience sample of twenty-three eight-grade students from a single educational institution (middle-school students). Consequently, the findings may not be generalizable to different educational settings, age, demographics, or levels of English language proficiency. Second, the study utilized a one-group pretest-posttest quasi-experimental framework with no control group. Although this design is suitable and practical for classroom-based Action Research, the lack of a comparison group limits the ability to attribute observed improvements solely to the instructional model.

Third, the intervention spanned a short six-week period. A more extended intervention may be necessary to provide additional evidence in terms of the sustained effects of AI-assisted instruction on writing development and long-term retention of organizational writing skills. Lastly, the study targeted specifically the organizational components of academic writing. Including paragraph structure, sentence cohesion and logical coherence. Other important aspects, such as critical thinking, grammar, vocabulary, punctuation, writing fluency, were not examined, but it remains for future research opportunities.

RESULTS

The general objective of this study was to determine how the use of ChatGPT as an AI-assisted tool, within a backward design approach, can support students in paragraph structure, sentence cohesion, and logical coherence. To establish a diagnostic baseline prior to the intervention, students' drafted paragraphs were evaluated through a 100-point rubric measuring the three-structural dimensions. Table 1 shows the statistics of the pre-assessment ($N = 23$) demonstrating the entry level of the students' writing proficiency

Table 1

Descriptive Statistics of the Pretest

N	Domain	Min	Max	M	SD
	Paragraph Structure	20.0	40.0	30.4	7.7
23	Sentence Cohesion	10.0	30.0	17.8	6.7
	Logical Coherence	10.0	30.0	18.3	6.5
	Total Score	**40.0**	**100.0**	**66.5**	**16.4**

The results presented in Table 1 ($N = 23$) indicate that students started the study with varying levels of writing proficiency across the different writing components. In the paragraph structure domain, participants recorded an initial mean score of *M* = 30.4 with *SD* = 7.7, spanning from Min = 20.0 to Max = 40.0 out 40 possible points. Conversely, learners demonstrated great difficulty in sentence cohesion and logical coherence. In sentence cohesion, a mean result of 17.8 (*M*) out of 30 points, *SD* = 6.7 and a minimum of 10.0 and a maximum of 30.0. In logical coherence, *M*= 18.3, *SD* = 6.5, *Min* = 10.0 and *Max* = 18.3 out of 30 points as well. Yielding an overall group mean Total score of M = 66.5 (*SD* = 16.4) out of 100 points.

Table 2

Descriptive Statistics of the Posttest

N	Domain	Min	Max	M	SD
	Paragraph Structure	20.0	40.0	33.5	5.7
23	Sentence Cohesion	20.0	30.0	26.5	4.9
	Logical Coherence	20.0	30.0	27.0	4.7
	Total Score	**70.0**	**100.0**	**87.0**	**11.5**

As presented in table 2, the administration of the post-test evaluation revealed a significant increase in all domains, but above all in sentence cohesion, and logical structure. The post-test results show that in the paragraph structure domain reached at M= 33.5 (SD = 5.7) out of 40 points, going from Min = 20.0 to Max = 33.5. Sentences cohesion and logical coherence revealed high-performance progress. Sentence cohesion with *M* = 26.5 (*SD* = 4.9), spanning from 20.0 to 30.0, and logical coherence shows a *M* = 27.0 (*SD* = 4.7), *Min* = 20.0 to *Max* = 30.0. out of 30 points each domain. Giving as a group Total score of M = 87.0 (*SD* = 11.5) out of 100 points.

Table 3

Description of the difference between the means

N	Pretest		Posttest		Difference between the means	
23	**M**	**SD**	**M**	**SD**	**MD**	**SD**
	66.5	16.4	87.0	11.5	20.4	7.7

As shown in Table 3, the results exhibit the overall score progression between the pretest and posttest. Prior to the intervention, participants obtained an overall group score of *M* = 87.0 (*SD* = 17.4). Once administered the six-week intervention plan, the posttest assessment results show a group mean score of *M* = 87.0 (*SD* = 11.5) out of 100 points. This increase represents an average group score gain of *MD* = 20.4 points (*SD* = 7.7), demonstrating a notable increase in the overall paragraph writing organization.

Table 4

Descriptive Statistics and Paired Samples t-Test for Rubric Domain Scores: Pre-test and Post-test (N = 23)

	PRETEST		**POSTTEST**						
Domain	**M**	**SD**	**M**	**SD**	**ΔM**	**T**	**df**	**p**	**Cohen's d**
Paragraph Structure	30.4	7.70	33.50	5.70	+3.0	3.10	22	.005	0.65
Sentence Cohesion	17.8	6.70	26.50	4.90	+8.7	12.11	22	< .001	2.53
Logical Coherence	18.3	6.50	27.00	4.70	+8.7	12.11	22	< .001	2.53
Total Score	**66.5 (16.4)**		**87.0 (11.5)**		**+20.4**	**12.77**	**22**	**< .001**	**2.66**

Note. *N* = 23. *M* = mean; *SD* = standard deviation; ΔM = post-test minus pre-test mean gain; t = paired-samples t-statistic; df = degrees of freedom; p = two-tailed significance value; Cohen's d interpreted as small = 0.20, medium = 0.50, large = 0.80+. Source: Author's elaboration (2026).

With the objective of determining how the use of ChatGPT as an AI-supported instructional tool, within a backward design approach, can support students in structuring paragraphs, improve coherence, and strengthen their use of liking devices through the writing process. To evaluate whether the observed improvements were statistically significant as well as to determine the impact of the intervention, paired-sample t-tests and Cohen's d effect sizes were calculated for each domain and for the overall composite score. Table 4 presents the descriptive and inferential changes comparing the pretest and posttest results.

In the paragraph structure domain, the group mean increased from *M* = 30.4 (*SD* = 7.70) in the pretest to *M* = 33.50 (*SD* = 5.70) in the posttest, resulting in a mean gain $\Delta M = +3.0$ points. To confirm this statistically, a pair-samples t-test was employed ratifying this improvement was significant ($t(22) = 3.10, p = .005$), with a Cohen's *d* of 0.65 indicating a moderate-to-large effect size.

Concerning sentence cohesion, students demonstrated a remarkable performance increase, moving from a pre-test mean of *M* = 17.8 (*SD* = 6.70) to a post-test mean of *M* = 26.50 (*SD* = 4.90). The point gain $\Delta M = +8.7$ was statistically considerable ($t(22) = 12.11, p < .001$) and yielded a large effect size ($d = 2.53$).

Regarding sentence coherence, participant scores improved from an initial mean of *M* = 18.3 (*SD* = 4.70) in the pretest to *M* = 27.00 (*SD* = 4.70) in the posttest, representing an average

gain of $\Delta M = +8.7$ points. Inferential testing showed that this gain again was statistically significant ($t(22) = 12.11, p < .001$) and yielded a large effect size ($d = 2.53$).

The overall performance rose from a pre-test mean of M = 66.5 (*SD* = 16.4) to a post-test mean of M = 87.0 (*SD* = 11.5), reflecting an average overall gain of $\Delta M = +20.4$ points. The difference was statistically important ($t(22) = 12.77, p < .001$), with a total effect size of d = 2.66, reflecting a strong positive impact on students' writing process.

The domain components that improved after intervention are sentence cohesion and logical coherence. Both subdimensions reveal domain level means of $\Delta M = +8.7$ points, advancing from baseline means of M = 17.8 and M = 18.3 to post-test means of M = 26.50 and M = 27.00. These gains represent a 48.9% increase in sentence cohesion and a 47.5% in logical coherence, both with large effect sizes ($d = 2.53, p < .001$). In contrast, paragraph structure shows a smaller mean gain of $\Delta M = +3.0$ points ($M = 30.4$ to $M = 33.50, d = 0.65$), however, all the three domains demonstrate mean gains.

Student Perception of AI-Supported Formative Feedback

To explore the students' perceptions of the use of AI during the intervention by using a pre-test and post-test and a survey with a Likert scale, participants completed an 18-item Likert scale survey ($1 =$ Totally Disagree to $5 =$ Totally Agree) upon completion of the 6-week intervention ($N = 23$). The instrument measured perceptions across two main domains: perceived impact on writing organization (items 1–9) and the pedagogical utility of the AI scaffolding (items 10–18).

As detailed in Table 4, overall student reception of the AI-supported instructional framework was highly positive, yielding an overall aggregate mean of $M = 4.45$ ($SD = 0.54$).

Table 4

Descriptive Statistics for Student Post-Intervention Perception Survey (N = 23)

Item #	Mean (M)	Standard Deviation (SD)	Agreement (%)
1	4.35	0.65	87.0%
2	4.48	0.51	95.7%
3	4.39	0.58	91.3%
4	4.57	0.51	95.7%
5	4.61	0.50	100.0%
6	4.30	0.63	87.0%
7	4.43	0.59	91.3%
8	4.35	0.57	91.3%
9	4.30	0.64	87.0%
10	4.52	0.51	95.7%
11	4.65	0.49	100.0%
12	4.57	0.51	95.7%

13	4.39	0.58	91.3%
14	4.52	0.51	95.7%
15	4.61	0.50	100.0%
16	4.43	0.59	91.3%
17	4.35	0.57	91.3%
18	4.61	0.50	100.0%
Overall Mean	4.45	0.55	92.4%

Note. $N = 23$. Measured on a 5-point Likert scale (1 = Totally Disagree to 5 = Totally Agree). Agreement (%) combines "Agree" and "Totally Agree" responses. Source: Author's elaboration (2026).

Within the writing organization domain, items related to *Sentence Cohesion* recorded the highest levels of student endorsement, particularly regarding '*replacing repetitive linkers'* (Item 5: $M = 4.61$, $SD = 0.50$, 100% agreement). Concerning AI scaffolding, students agreed that ChatGPT worked as a collaborative tutor rather than an automated substitute, endorsing the assertions that AI supported their drafting process without giving automated-generated text for them (Item 11: $M = 4.65$, $SD = 0.49$; Item 15: $M = 4.61$, $SD = 0.50$). Furthermore, 100% of participants agreed that having teacher instruction with AI support maximized learning effectiveness (Item 18: $M = 4.61$, $SD = 0.50$). Giving an overall mean of M = 4.45 (SD = 0.55) with a global agreement of 92.4%.

DISCUSSION

The primary specific objective of this research was to determine how the use of ChatGPT as an AI-supported instructional tool, within a backward design approach, can support students in paragraph structure, sentence cohesion, logical coherence. Therefore, the findings of this action research study demonstrate a significant improvement in eight-graders' paragraph organization. Among the gains observed, it is shown a mean gain ($\Delta M = +20.4$) and a large effect size ($d = 2.66$) exceeding standard benchmarks for other field interventions, where and effect size above $d = 0.80$ are interpreted as large (Plonsky & Oswald, 2014). Participants showed the greatest growth in sentences cohesion ($d = 2.53$) and logical coherence ($d = 2.53$). Although paragraph Structure showed a smaller effect size ($d = 0.65$), this is due to the high pre-test performance ($M = 30.4$ out of 40), reflecting ceiling effects where students already had basic knowledge of topic sentence placement prior to intervention.

Comparing the results obtained, these quantitative and structural gains align with recent empirical findings by Stitou and Fatmi (2026) which investigated the effects of AI-assisted formative feedback on the writing performance among secondary learners. Their study demonstrated statistically significant improvements in overall writing performance ($M_{pre} = 11.85, SD = 2.41$ vs. $M_{post} = 15.62, SD = 1.98; t(44) = 11.24, p < .001$), with a large effect

size ($d = 1.68$). These findings confirm that the use of AI-assistant tool (ChatGPT) contributed to a measurable improvement in learners' writing organization specifically in content organization, sentence cohesion, and sentence coherence.

The results of this study are also consistent with other research on AI-assisted writing tools. Recent studies on writing evaluation using AI have shown that these tools can enhance students' writing skills across different components, sentence structure, sentence cohesion, paragraph coherence, including grammar and vocabulary (Ekizoğlu1 & Demir, 2025). According to this study, the experimental group showed grater improvement in overall writing scores, demonstrating that the use of AI as a scaffolding tool rather than as a text provider can significantly increase their writing process.

The second specific of this study was to explore students' perceptions and reactions to the use of AI during the intervention by using a pretest, posttest, a survey with a Likert scale, and teacher field notes. Based on the post-intervention Likert scale survey, students demonstrated positive attitudes towards the use of ChatGPT as a scaffolding tool yielding a mean score of $M = 4.45/5.00$ (92.4 % overall positive agreement). Students indicated that ChatGPT served as an interactive, Socratic guide that facilitated revision, and gave immediate feedback to help them identify organizational gaps. These findings align directly with Song and Song (2023), whose qualitative data revealed that 88% of students perceived ChatGPT helpful, noting that the immediate AI assistance increased their writing motivation ($M = 20.06/25.00, p = 0.001$). Furthermore, these perception data triangulate with qualitative field notes, which documented a noticeable reduction of writing anxiety, and an increase in drafting confidence. Participants also expressed that integrating AI feedback within teacher-led, Background design structured maximized their writing progress.

Some limitations should be considered regarding the study. First, the sample size ($N = 23$) and single-classroom action research design limit broad generalizability to several contexts. Second, the lack of a control group prevents complete isolation of confounding variables such as general cognitive maturation. Future studies should employ a quasi-experimental pretest-posttest with a control group design and incorporate delayed post-tests to assess the long-term retention of structural writing gains.

CONCLUSIONS

This research study focused on addressing structural writing challenges of writing organization among eight-grade EFL students by incorporating a six-week intervention using a Backward Design supported by ChatGPT as an AI-scaffolded tool for formative feedback. The empirical findings provide strong quantitative and qualitative evidence that structured AI scaffolding can significantly improve writing organization. The post-intervention evaluation demonstrated a statistically significant improvement in overall paragraph writing performance,

with the class mean increasing from a baseline of $M = 66.5$ ($SD = 16.4$) to $M = 87.0$ ($SD = 11.5$), $t(22) = 12.77, p < .001$. The resulting effect size ($d = 2.66$) confirms a highly potent pedagogical impact, confirming the integration of automated formative guidance within middle school contexts.

The domain analysis shows that the pedagogical intervention was most effective in resolving relational and cohesive writing deficits of the overall writing. *Sentence Cohesion* and *Logical Coherence* exhibit the largest parametric gains ($\Delta M = +8.7, d = 2.53$), confirming that real-time, iterative feedback can help learners in eliminating redundant linkers, diversifying transitional phrases, and keeping coherence through the paragraph construction. Furthermore, the instructional framework successfully increased overall performance tiers, eliminating the *Beginning* achievement category (26.1% to 0%), advancing 30.4% of students to a proficient level, and enabling more than half of the sample (52.2%) to reach the advanced benchmark, resulting in an overall proficiency of 82.6%.

The student perception survey ($M = 4.45, SD = 0.54$) supports these empirical improvements by revealing high levels of learner's motivation and positive attitude towards the use of AI Assistance. Participants overwhelmingly perceived ChatGPT as an interactive, Socratic guide rather than a simple text generator, emphasizing that the tool encouraged active self-correction and revision maintaining their critical thinking. Crucially, unanimous student agreement (100%) concerning the combined delivery of teacher instruction and AI support highlights that artificial intelligence (AI) achieves its full efficacy when embedded within an intentional, teacher-led curriculum rather than implemented as a separate technology. Ultimately, this research offers an expandable model for middle school educators, demonstrating that ethical and structured AI integration enhances foundational writing skills while fostering autonomous language learning habits.

REFERENCES

Agustini, M., Ramendra, D., & Adnyani, N. (2025). Analysis of Coherence and Cohesion in Paragraphs Writing of EFL Students. *Journal of Educational Study, 5*(1), 23-24. doi:10.36663/joes.v5i1.966

Baylon, M. (2023). Instructional scaffolding strategies for teaching writing to junior high school students of a general public school. *Asian Journal of Language, Literature and Culture Studies, 6*(3), 302-314.

Burns, A. (2010). *Doing Action Research in English Language Teaching: A Guide for Practitioners.* Routledge.

Creswell, J., & Creswell, J. (2023). *Research design: qualitative, quantitative, and mixed methods approaches* (6th edition ed.). SAGE Publications.

East, M. (2009). Evaluating the reliability of analytic rubrics for assessing second language writing. *Assessing Writing, 14*(2), 88-105. doi:10.1016/j.asw.2009.04.001

Ekizoğlu1, M., & Demir, A. (2025). The role of AI assisted writing feedback in developing secondary students writing skills. *Discover Education, 4*(1), 454. doi:10.1007/s44217-025-00919-3

Graham, S., Cao, Y., Kim, Y.-S., Lee, J., Tate, T., Collins, P., . . . Olson, H. (2024). Effective writing instruction for students in grades 6 to 12: a best evidence meta-analysis. *Reading and Writing, 38*, 1–46. doi:10.1007/s11145-024-10539-2

Halliday, M., & Hasan, R. (1976). *Cohesion in English.* London: Longman.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Daryna, F., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*. doi:10.1016/j.lindif.2023.102274

Kim, J., Yu, S., Detrick, R., & Li, N. (2025). Exploring students' perspectives on Generative AI-assisted academic writing. *Education and Information Technologies, 30*(1), 1265–1300. doi:10.1007/s10639-024-12878-7

Li, M., & Wilson, J. (2025). AI-Integrated scaffolding to enhance agency and creativity in K-12 english language learners: A systematic review. *Information, 16*(7), 519. doi:10.3390/info16070519

Plonsky, L., & Oswald, F. (2014). How Big Is “Big”? Interpreting Effect Sizes in L2 Research. *Language Learning, 64*(4), 878-912. doi:10.1111/lang.12079

Riswanto, R. (2021). Teaching academic writing: Enhancing logical coherence and thematic progression in EFL classrooms. *International Journal of Instruction, 14*(3), 412-428.

Sarmiento, C., Truckenmiller, A., Cho, E., & Wang, H. (2024). Academic language use in middle school informational writing. *British Journal of Educational Psychology, 95*, 385-389. doi:10.1111/bjep.12724

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158. doi:10.1093/applin/11.2.129

Shakra, Z. (2023). Feedback as Scaffolded Instruction in the Assessment of L2. *European Scientific Journal, 19*(19), 57. doi:10.19044/esj.2023.v19n29p57

Siekmann, L., Parr, J., & Busse, V. (2022). Structure and coherence as challenges in composition: A study of assessing less proficient EFL writers' text quality. *Assessing Writing, 54*, 1-13. doi:10.1016/j.asw.2022.100672

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology, 14*. doi:10.3389/fpsyg.2023.1260843

Stitou, H., & Fatmi, H. (2026). The effect of generative AI-assisted feedback on EFL students' writing performance. *International Journal of Linguistics and Translation Studies, 7*(3), 25–38. doi:10.36892/ijlts.v7i3.739

Utami, A., & Bram, B. (2023). Backward design Implementation in English as a foreign language (EFL) context. *Journal of Education and Teaching, 4*(1), 110-118. doi:10.51454/jet.v4i1.211

Wakerkwa, D. (2024). Grammatical cohesive features in Indonesian EFL students' argumentative essays. *Multidisciplinary Research Journal, 02*(01), 56-63. doi:10.70210/amrj.v2i01.63

Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: a randomized controlled trial. *Frontiers in Psychology, 14*, 1249991. doi:10.3389/fpsyg.2023.1249991

Wiggins, G., & McTighe, J. (2005). *Understanding by Design.* Association for Supervision and Curriculum Development.

Zhang, K. (2025). Enhancing critical writing through AI feedback: A randomized control study. *Behavioral Sciences, 15*(5), 600. doi:10.3390/bs15050600

Zhang, W. (2026). Research on teaching strategies of the "Practical reading and communication" unit in primary school chinese based on backward design. *Journal of Education and Educational Policy Studies, 4*(2), 11-23. doi:10.54254/3049-7248/2026.31916