

## CHATGPT AS A FORMATIVE FEEDBACK TOOL: IMPROVING NARRATIVE ESSAY WRITING AMONG EFL STUDENTS

Pariyanto<sup>1</sup>, Novalita Fransisca Tungka<sup>2\*</sup>

<sup>1</sup>[pariyanto@untag-sby.ac.id](mailto:pariyanto@untag-sby.ac.id), <sup>2\*</sup>[novalita@unsimar.ac.id](mailto:novalita@unsimar.ac.id)

<sup>1</sup>Universitas 17 Agustus 1945 Surabaya, <sup>2</sup>Universitas Sintuwu Maroso Poso

Received: 27<sup>th</sup> December 2024

Revised: 17<sup>th</sup> June 2025

Accepted: 28<sup>th</sup> June 2025

**ABSTRACT** As Artificial Intelligence (AI) tools become increasingly accessible in educational contexts, their potential to support second language writing instruction warrants closer examination. This study investigates the effectiveness of formative feedback provided by ChatGPT, an AI-based language model, in enhancing the narrative writing skills of English as a Foreign Language (EFL) learners. Sixteen undergraduate students from an Indonesian university participated in the study. Each student composed a narrative essay based on a personal experience and received individualized feedback from ChatGPT using a standardized prompt. They subsequently revised their work based on this feedback. Pre- and post-revision essays were evaluated by two trained human raters using a standardized narrative writing rubric. Inter-rater reliability was confirmed with a high Cronbach's alpha score of .91. Statistical analysis using paired-sample t-tests revealed significant improvements in overall writing quality, with the greatest gains observed in style, voice, and coherence. Qualitative data from student reflections further indicated positive perceptions of ChatGPT's feedback, citing its clarity, detail, and ease of use. While some students noted the tool's limitations in addressing cultural or contextual nuances, overall results suggest that AI-generated feedback can serve as an effective supplement to traditional instruction. The study highlights ChatGPT's potential to enhance writing development and learner autonomy, particularly in resource-constrained EFL classrooms.

**Keywords:** Artificial Intelligence (AI), ChatGPT, formative feedback, narrative writing, EFL learners

### INTRODUCTION

The rapid advancement of digital technologies has profoundly reshaped educational practices across the globe. In recent years, the integration of technology into classroom instruction has moved beyond the use of basic digital tools and multimedia presentations to include sophisticated applications such as Artificial Intelligence (AI), Virtual Reality (VR), and Augmented Reality (AR) (Chadha, 2024; Kara et al., 2024; Pariyanto et al., 2025). Among these emerging technologies, AI has shown remarkable potential to enhance teaching and learning by offering scalable, personalized, and data-driven instructional support. Within the domain of language education, AI tools are beginning to play a central role in helping learners develop core linguistic competencies, including writing skills (Dwivedi et al., 2023; Godwin-Jones, 2024).

One of the most influential developments in this field is ChatGPT, an advanced AI language model developed by OpenAI. Trained on vast corpora of text data and fine-tuned to generate human-like responses, ChatGPT has attracted attention not only for its conversational capabilities but also for its capacity to deliver instant, formative feedback on written texts. This feedback is typically comprehensive, covering aspects such as grammar, coherence, organization, style, and tone (OpenAI, 2022). Such capabilities position ChatGPT as a potentially transformative tool in the teaching of English as a Foreign Language (EFL), particularly in contexts where learners have limited access to individualized support from instructors (Mahapatra, 2024).

In the realm of second language writing instruction, feedback is widely recognized as one of the most effective pedagogical tools (Jiang & Hyland, 2025; Mao, Lee, & Li, 2024). Formative feedback, in particular, plays a crucial role in helping learners identify errors, reflect on their writing choices, and revise their texts for improvement (Yan, He, & Sheng, 2024). However, providing high-quality, individualized feedback is a time-intensive task, especially in classrooms with high student-teacher ratios, a common feature of many EFL learning environments, including in Indonesia (Zebua & Katemba, 2024). Consequently, EFL

instructors often struggle to provide timely and comprehensive feedback on students' writing, resulting in limited opportunities for learners to engage in meaningful revision practices. Here, AI tools like ChatGPT present a viable solution by automating the feedback process and making it available on demand (Banihashem et al., 2024).

Research into AI-supported writing instruction has grown considerably in recent years. Several studies have reported positive outcomes associated with the use of AI-based tools such as Grammarly, Quillbot, and ChatGPT in improving grammatical accuracy, lexical range, and coherence (Yan, 2024; Polakova & Ivenz, 2024). These tools have also been associated with increased learner motivation and autonomy, as students are able to engage with feedback iteratively and at their own pace (Mahapatra, 2024). However, the majority of this research has focused on expository or academic writing tasks, such as essays, reports, and summaries. Comparatively little attention has been paid to how AI tools perform when applied to narrative writing, which involves different challenges and evaluative criteria.

Narrative writing, as a genre, is fundamentally distinct from expository writing. While the latter emphasizes logical structure, clarity of argument, and evidential support, the former values creativity, emotional engagement, character development, and storytelling techniques (Ngoi, Tan, Alias, & Mat, 2024). In EFL contexts, narrative writing tasks are often used to develop learners' ability to express themselves, experiment with language, and communicate personal or imaginative experiences (Saaty, 2023). As such, successful narrative writing depends not only on linguistic accuracy but also on style, voice, and coherence, dimensions that require more nuanced and context-sensitive feedback. Whether AI tools like ChatGPT are capable of recognizing and effectively responding to these dimensions is a question that remains largely unanswered in the current literature.

Adding to this gap, many existing studies on AI feedback tools are based in high-resource educational settings, often in Western countries, where access to digital infrastructure and teacher training is relatively advanced (Dwivedi et al., 2023; Godwin-Jones, 2024). However, in developing contexts such as Indonesia, the integration of AI tools into language instruction is still in its early stages. EFL learners in these contexts frequently encounter challenges such as outdated curricula, limited instructional support, and restricted access to native or proficient speakers of English (Zebua & Katemba, 2024). As such, examining the efficacy of AI tools in these environments is critical not only for academic understanding but also for practical implementation in under-resourced educational systems.

Furthermore, while the technological capabilities of ChatGPT are well-documented, learners' perceptions and experiences with using AI tools for writing support are less frequently explored (Yan, 2024). Understanding how students interpret, respond to, and act upon the feedback they receive is essential for assessing the pedagogical value of such tools. Feedback that is technically accurate but difficult for learners to understand or implement may have limited effectiveness. Conversely, feedback that learners perceive as clear, supportive, and actionable is more likely to lead to meaningful revision and improvement. Therefore, investigating student perceptions alongside quantitative measures of writing performance provides a more holistic view of AI's role in supporting language learning.

Against this backdrop, the current study aims to address three key gaps in the existing research: (1) the limited focus on narrative writing in AI feedback studies, (2) the scarcity of research in EFL contexts in the Global South, particularly Indonesia, and (3) the need to explore both performance outcomes and learner perceptions. The study investigates how formative feedback generated by ChatGPT affects the narrative essay writing of undergraduate EFL learners. Specifically, it examines whether the use of ChatGPT leads to measurable improvements in core writing dimensions such as style, voice, creativity, and coherence, and how learners perceive the usefulness, clarity, and impact of the feedback they receive.

The study involved sixteen undergraduate students majoring in English Literature at an Indonesian university. Each participant wrote a narrative essay based on a personal memory, submitted it for feedback through ChatGPT, and revised it accordingly. The pre- and post-feedback essays were assessed using a standardized rubric by two trained human raters to determine the impact of the AI intervention. Additionally, interviews and reflective comments were collected to capture the learners' subjective experiences and perceptions.

In doing so, this study contributes to the growing body of scholarship on AI-assisted language learning and the pedagogical integration of emerging technologies in resource-constrained settings. It offers empirical evidence on how ChatGPT can support not only language accuracy but also higher-order writing skills such as creativity and voice in EFL narrative writing. Moreover, the study highlights practical implications for language educators, curriculum designers, and policy-makers seeking to leverage AI tools to enhance teaching effectiveness and learner outcomes.

By addressing both the quantitative improvements in writing performance and the qualitative insights from learners, this research provides a balanced and contextually grounded evaluation of ChatGPT's utility in EFL instruction. Ultimately, it encourages a nuanced and critical approach to integrating AI into language education, one that recognizes both its potential and limitations while striving to make meaningful improvements in learners' writing development.

## **LITERATURE REVIEW**

### **Feedback and Language Acquisition**

Feedback plays a central role in second language acquisition (SLA), especially in writing development. In language learning contexts, feedback is broadly categorized into formative and summative types. Formative feedback, in particular, is designed to guide learners during the learning process, helping them reflect, revise, and improve their output (Yan, He, & Sheng, 2024). According to Fisher, Brotto, Lim, & Southam (2025), effective formative feedback should be clear, timely, specific, and geared toward promoting self-regulated learning. In writing pedagogy, such feedback fosters improvement in core dimensions like grammatical accuracy, organizational structure, coherence, and idea development (Jiang & Hyland, 2025).

Writing is a complex cognitive and linguistic task requiring sustained attention to language form, structure, content, and audience. In EFL (English as a Foreign Language) settings, the difficulty is compounded by limited input, practice opportunities, and feedback availability. Learners rely heavily on teachers' written comments to improve their texts. However, as Jiang & Hyland (2025) and Mao, Lee, & Li (2024) observed, the effectiveness of teacher feedback depends not only on its content but also on the frequency, clarity, and appropriateness of delivery. Teachers, especially in large classes, often struggle to provide individualized and detailed feedback consistently. This reality has sparked interest in alternative or complementary feedback systems, particularly those facilitated by digital technologies.

### **Challenges in EFL Writing Feedback**

In EFL classrooms worldwide, including Indonesia, traditional feedback models face numerous challenges. Limited instructional time, large class sizes, and uneven teacher training lead to inconsistencies in the quality and quantity of feedback provided to learners (Joe & Mtsi, 2024). Moreover, students frequently express frustration over vague or overly general comments, such as "be more clear" or "improve grammar," which fail to provide actionable steps for improvement (Yunita & Kusuma, 2022). These shortcomings limit the formative function of feedback, reducing its potential to promote learner autonomy and linguistic development.

Adding to this complexity is the emotional dimension of writing feedback. Learners often interpret teacher comments as judgmental or demotivating, particularly when delivered without encouragement or specific guidance (Jiang & Hyland, 2025; Mao, Lee, & Li, 2024). In many cases, students fail to revise or understand the corrections, treating feedback as a final grade rather than a developmental opportunity (Afifi, Rahimi, & Wilson, 2023). These pedagogical and affective barriers underscore the need for scalable, accessible, and non-threatening feedback mechanisms, an area where Artificial Intelligence (AI) may offer meaningful contributions.

### **AI-Supported Feedback Tools in Writing Instruction**

The rise of AI in education has introduced new possibilities for personalized and scalable feedback systems. AI-powered writing assistants, such as Grammarly, Quillbot, and more recently, ChatGPT, offer real-time suggestions on grammar, punctuation, vocabulary, and organization. These tools operate on natural language processing (NLP) algorithms trained on massive text corpora, enabling them to identify linguistic patterns and generate human-like responses (Huang, 2024; OpenAI, 2022).

Among these, ChatGPT has received particular attention for its adaptability, interactivity, and fluency in providing feedback. As an open-domain chatbot, ChatGPT can analyze student writing and generate structured, detailed responses that mimic the scaffolding provided by human instructors. Unlike older grammar-checking tools, ChatGPT is capable of addressing higher-order concerns such as coherence, logical progression, and narrative flow. Mahapatra (2024) reported that ESL students who received feedback from ChatGPT demonstrated notable improvements in grammar, organization, and lexical variety. Similarly, Polakova and Ivenz (2024) found that ChatGPT feedback enhanced the overall coherence and structure of EFL students' argumentative essays.

While much of the early AI-feedback research has focused on expository or argumentative writing, studies are beginning to explore broader applications. Zebua and Katemba (2024) investigated learners' perceptions of using ChatGPT for writing tasks and found that students appreciated its accessible interface and comprehensive feedback. However, they also noted some limitations, particularly the AI's lack of cultural awareness and sensitivity to rhetorical conventions.

### **Theoretical Underpinnings: SLA and Feedback**

The use of ChatGPT for formative feedback is theoretically grounded in established SLA frameworks. Krashen's Input Hypothesis (1982) posits that learners acquire language most effectively when they are exposed to input that is slightly beyond their current proficiency level ( $i+1$ ). ChatGPT, when properly prompted, can provide input that matches this criterion by identifying learners' current writing issues and offering linguistically rich alternatives. This makes AI feedback a valuable source of modified, comprehensible input, especially in contexts where teacher-student interaction is limited.

Hamzah (2023) suggests that interaction further supports the role of feedback in SLA, emphasizing the importance of negotiation of meaning in language development. While AI does not offer authentic social interaction, it can simulate some aspects of dialogic feedback through iterative engagement. For example, learners can ask ChatGPT for clarification, elaboration, or alternative phrasing, creating a quasi-interactive environment that encourages deeper cognitive engagement with language.

Another important theoretical foundation is Ekanayaka & Ellis's (2024) work on corrective feedback. Ekanayaka & Ellis (2024) emphasize that focused and timely corrective feedback enhances learners' metalinguistic awareness and contributes to their internalization of language rules. AI tools like ChatGPT deliver this feedback instantly and can be customized



to target specific linguistic areas, such as tense usage, voice, or cohesion. In addition, learners can control the pace of engagement, revisiting the AI's suggestions multiple times, which supports autonomous learning and reinforces linguistic accuracy.

### **ChatGPT in the EFL Writing Classroom**

ChatGPT's relevance in writing instruction goes beyond its technical sophistication. One of its most valued features is its scalability, which makes it suitable for settings with limited instructional resources. Unlike traditional one-on-one writing tutorials, which are time-consuming and often inaccessible to many students, ChatGPT provides round-the-clock assistance. Students can receive detailed explanations of errors, suggestions for improvement, and sample rewrites, all within seconds. This immediacy and accessibility are particularly beneficial in EFL classrooms in Southeast Asia, where classroom sizes are large and access to native-speaking instructors is scarce (Ekizoğlu & Demir, 2025).

Furthermore, ChatGPT can provide feedback in a non-threatening and anonymous environment, which some students find preferable to teacher or peer evaluations. Learners may be more willing to take risks, revise their work, and explore alternative phrasing when working with an AI system that offers neutral, judgment-free suggestions (Yan, 2024). However, this convenience comes with trade-offs. AI feedback may not be sensitive to discourse-level nuance, cultural context, or genre-specific conventions. For example, while ChatGPT might suggest grammatically correct alternatives, it may overlook subtleties in tone, audience expectation, or stylistic coherence in narrative writing.

### **Gaps in the Literature: Narrative Writing and AI Feedback**

Despite the growing interest in AI-assisted feedback, research on its effectiveness in narrative writing remains limited. Narrative writing requires a unique set of competencies, including voice, emotional resonance, character development, and narrative arc. These elements are more subjective and harder to quantify than grammatical correctness or paragraph organization. Most AI tools, including ChatGPT, are optimized for syntactic and semantic analysis but may struggle to evaluate storytelling elements effectively (Ibrahim & Kirkpatrick, 2024).

Narrative writing in EFL contexts is often used to foster creativity, personal expression, and engagement with the target language. It provides students with opportunities to use language in authentic and meaningful ways, thereby strengthening their linguistic competence and confidence. However, because narrative tasks are more open-ended and varied in structure, they present unique challenges for automated feedback systems. This study responds to that gap by examining how ChatGPT performs as a feedback tool in the context of EFL students' narrative writing. Specifically, it investigates whether ChatGPT can enhance core narrative dimensions such as style, voice, coherence, and creativity, and how students perceive the AI's role in their writing development.

## **RESEARCH METHODS**

### **Research Type and Design**

This study employed a quantitative quasi-experimental design with a one-group pre-test and post-test structure. It was supported by a descriptive qualitative component to provide insight into students' perceptions of the feedback process. The main aim was to measure the impact of AI-generated formative feedback from ChatGPT on the narrative writing performance of EFL learners. This mixed-methods approach allowed the researchers to not only assess the statistical significance of improvement in writing scores but also to understand the learners' subjective experiences and responses to the AI tool.

## Population and Sample

The population of this study consisted of undergraduate students enrolled in the English Literature program at Universitas 17 Agustus 1945 Surabaya, Indonesia. The sample was selected using purposive sampling, targeting students who had completed foundational courses in paragraph and essay writing. A total of 16 students (5 females and 11 males), all in their third or fourth year of study, voluntarily participated in the study. Their TOEFL prediction scores ranged from 400 to 500, indicating intermediate-level English proficiency. All participants had previous exposure to academic writing but had not previously received AI-generated feedback.

## Data Collection Techniques and Instruments

Data collection was carried out in two phases:

### *Writing Task and Feedback Cycle*

Each participant was assigned a narrative writing task, where they were asked to write a 300–500-word essay on the topic “a childhood memory”. This prompt was chosen to allow students to engage with personal storytelling and demonstrate their ability to use narrative structures. After completing the initial draft (pre-test), students submitted their essays to ChatGPT (GPT-4) via a standard prompt: *"Please review this narrative essay and offer formative feedback, with particular attention to grammar, coherence, creativity, and overall organization. Include specific suggestions for improving each of these areas."* The feedback generated by ChatGPT was then reviewed by each participant, who subsequently revised their essays independently based on the feedback provided (post-test version).

### *Qualitative Feedback Collection*

After submitting the revised essays, participants completed a brief open-ended questionnaire reflecting on their experience using ChatGPT feedback. The questions focused on feedback clarity, usefulness, perceived improvements, and challenges. These responses were used to enrich the interpretation of the quantitative findings.

## Measurement and Operational Definitions of Variables

The primary dependent variable in this study was narrative writing quality, operationally defined as the score achieved on a standardized narrative writing rubric. The rubric was adapted from the City University of New York (CUNY) Narrative Essay Rubric, and it consisted of the following six components:

- Ideas – clarity, focus, and depth of the narrative theme
- Style – lexical choice, sentence variation, tone
- Voice – personal engagement and distinctiveness in writing
- Creativity – originality, vividness, and descriptive detail
- Transitions – logical flow between paragraphs and ideas
- Conventions – grammar, spelling, punctuation, and mechanics

Each component was rated on a 5-point Likert scale (1 = Poor, 5 = Excellent), resulting in a maximum possible score of 30 points per essay.

The independent variable was the ChatGPT-provided formative feedback, defined as text-based commentary generated in response to each student's initial narrative draft, based on the provided standardized prompt.

## Scoring and Inter-Rater Reliability

To ensure the objectivity and consistency of assessment, both the pre-test and post-test essays were independently evaluated by two trained human raters, faculty members experienced in teaching academic and narrative writing. The raters used the rubric to assess each essay without knowing whether it was a pre- or post-intervention version.

Inter-rater reliability was established by calculating Cronbach's alpha, yielding a value of .91, which indicates a high level of agreement and reliability between the two raters.

### Data Analysis Procedures

The collected data were first compiled and organized using Microsoft Excel and then analyzed using IBM SPSS Statistics 26. The main statistical test employed was a paired-sample t-test, which compared the pre-feedback and post-feedback mean scores of the same group to determine whether the revisions based on ChatGPT feedback led to statistically significant improvements in writing quality.

Descriptive statistics (mean, standard deviation) were reported for each rubric category, and inferential statistics were used to determine the significance of differences between pre- and post-test results. A significance threshold of  $p < .05$  was adopted.

Additionally, qualitative responses from the participant questionnaires were analyzed using thematic analysis, with recurring themes such as clarity, usefulness, motivation, and perceived limitations coded and summarized to complement the statistical findings.

## FINDINGS AND DISCUSSION

### Findings

#### *Quantitative Results*

The analysis of students' narrative essays before and after receiving formative feedback from ChatGPT revealed statistically significant improvements in writing quality. Using a paired-sample t-test, the mean total score increased from 69.69 in the pre-feedback stage to 74.75 after students revised their essays based on ChatGPT's suggestions. This improvement was statistically significant at  $p < .001$ , indicating a strong effect of the AI-generated feedback on students' performance.

*Table 1. Pre-and Post-Feedback Essay Scores*

Participant	Pretest	Post-test
1	51	53
2	64	72
3	67	74
4	87	89
5	68	74
6	76	80
7	74	82
8	71	75
9	85	87
10	67	74
11	58	68
12	77	81
13	77	79
14	73	76
15	66	71
16	54	61
	69,69	74,75

*Table 2. Mean Scores and Improvement*

Criteria	Pre-Feedback Mean	Post-Feedback Mean	Improvement
Ideas	15.06	15.63	+0.57
Style	13.19	14.56	+1.37

Criteria	Pre-Feedback Mean	Post-Feedback Mean	Improvement
Voice	14.56	16.00	+1.44
Creativity	6.19	6.63	+0.44
Transitions	3.25	3.63	+0.38
Conventions	3.00	3.44	+0.44

Table 3. Paired-Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Pretest - Posttest	-5.06	2.56	.64	-6.43	-3.69	-7.88	15	.000

A more detailed breakdown of the rubric categories shows that the most substantial improvements occurred in Voice (mean gain = +1.44), reflecting enhanced personal expression and engagement, Style (+1.37), indicating improved lexical choices, tone, and sentence variation.

Other areas showed moderate but positive gains in Creativity that improved by +0.44, Conventions (grammar, mechanics, punctuation) by +0.44, and Transitions (logical flow and paragraphing) by +0.38.

The overall mean difference of 5.06 points between pre- and post-test scores was supported by a t-value of -7.88 (df = 15, p = .000). The 95% confidence interval for the difference (ranging from -6.43 to -3.69) further confirms that the observed improvements were unlikely due to chance, underscoring the effectiveness of ChatGPT’s feedback in improving various aspects of narrative writing.

These findings suggest that ChatGPT was particularly effective in supporting higher-order writing skills such as voice and style, which are critical in narrative composition, while also offering support in grammar and structural coherence.

### Qualitative Insights

To complement the quantitative data, students’ perceptions were gathered through post-intervention open-ended reflections. The responses overwhelmingly revealed a positive attitude toward the feedback provided by ChatGPT.

Most students described the feedback as clear, specific, and actionable, allowing them to understand their weaknesses and revise their essays with confidence. One participant remarked, “It was very easy to understand because ChatGPT gives complete information about what I lack in my writing.”

Another noted, “It helped me fix the paragraph and structure of my essay.”

Many participants reported feeling more motivated and autonomous in their writing process, stating that the AI tool helped them revise their work without the fear of judgment. They appreciated how the feedback was direct yet supportive, guiding them to improve both surface-level accuracy and deeper narrative structure.

However, a few students also pointed out limitations of the AI-generated feedback. These included a lack of context-specific suggestions, particularly regarding cultural references or personal storytelling nuances that the tool did not fully understand. Some noted that while ChatGPT helped with grammar and organization, it was less helpful in refining the emotional tone or creativity of their stories.



Despite these caveats, the overall impression was that ChatGPT served as a useful, efficient, and supportive writing companion, especially in a context where access to individualized instructor feedback may be limited.

## **Discussion**

The results of this study reinforce the growing body of literature affirming the potential of artificial intelligence, particularly ChatGPT, as a valuable tool for formative feedback in English as a Foreign Language (EFL) writing classrooms. The statistically significant improvements observed across all assessed criteria, especially in style and voice, demonstrate that AI-generated feedback can effectively support EFL learners' ability to express themselves with greater clarity, creativity, and engagement in narrative writing.

### ***Supporting Expressive and Creative Writing***

Narrative writing is a genre that demands more than grammatical accuracy; it relies on the writer's ability to evoke emotion, build compelling characters, and maintain a coherent and engaging storyline. The notable improvements in voice (+1.44) and style (+1.37) suggest that ChatGPT is not merely functioning as a grammar checker but is providing feedback that resonates with learners at a deeper level of authorial control and expression.

These results align with previous findings by Mahapatra (2024) and Polakova & Ivenz (2024), who observed similar enhancements in learners' stylistic development and coherence after interacting with ChatGPT. These studies, and the present one, indicate that students are capable of leveraging AI-generated suggestions to reflect on their narrative choices and revise their work in meaningful ways. It appears that the natural, conversational tone of ChatGPT's feedback plays a key role in making such revisions approachable and constructive.

### ***ChatGPT as a Solution to Instructional Constraints***

The integration of ChatGPT also addresses longstanding structural challenges in language instruction, such as large class sizes, time limitations, and unequal access to individualized feedback. These issues are especially prevalent in developing EFL contexts, including Indonesia, where teacher workload often prohibits timely and detailed responses to each student's writing (Zebua & Katemba, 2024).

The fact that ChatGPT provides instant, context-relevant feedback allows students to engage in an iterative process of revision and reflection, practices that are often recommended in writing pedagogy but not always achievable in time-constrained classrooms (Jiang & Hyland, 2025; Mao, Lee, & Li, 2024). In this sense, ChatGPT complements traditional instruction by democratizing access to feedback, giving learners autonomy to revise and experiment without waiting for teacher input.

This finding corresponds with Steinert et al.'s (2024) formative assessment model, which stresses the importance of timely, actionable feedback in enabling learners to close the gap between current and desired performance. The AI's immediacy also supports self-regulated learning, as described by Fisher, Brotto, Lim, & Southam (2025), by allowing students to set goals, monitor progress, and revise texts independently.

### ***Theoretical Implications: Aligning with SLA Models***

The findings of this study are further supported by well-established Second Language Acquisition (SLA) theories, which emphasize the importance of input, interaction, and feedback in the learning process.

According to Krashen's Input Hypothesis (1982), language acquisition occurs most effectively when learners are exposed to comprehensible input that is just beyond their current proficiency level ( $i+1$ ). ChatGPT's feedback serves this purpose by offering scaffolded input,

highlighting errors while modeling improved structures and vocabulary that are slightly more advanced than the learners' original usage. Unlike static grammar tools, ChatGPT's language is adaptive and contextualized, which increases the likelihood of it being understood and internalized.

Hamzah (2023) posits that interaction and the negotiation of meaning play a key role in promoting language development. While ChatGPT does not replicate true dialogic negotiation in the human sense, its conversational interface and ability to simulate clarification, elaboration, and recasting can mimic some aspects of interactive learning. Students can prompt ChatGPT with follow-up questions, request alternative phrasing, or ask for further explanation, creating a quasi-interactive environment that promotes deeper cognitive processing of language.

From the perspective of Ekanayaka & Ellis's (2024) theory of corrective feedback, ChatGPT performs the dual function of both direct correction and metalinguistic explanation. For instance, rather than simply identifying an error, the tool often explains the nature of the error and provides examples for improvement. This kind of elaborated feedback is associated with higher learning gains in SLA research, as it helps learners not only fix errors but also understand the underlying linguistic rules.

### ***Interpreting the Modest Gains in Some Categories***

While the overall gains in writing quality were significant, some areas, such as transitions (+0.38) and conventions (+0.44), showed modest improvements relative to style and voice. Several explanations can be proposed for this trend.

First, conventions, such as grammar, punctuation, and spelling, may have already been relatively well-developed among the participants due to their prior training in foundational writing courses. As such, there was less room for noticeable improvement. Second, ChatGPT's ability to detect and explain conventional errors may vary depending on how the prompt is phrased or the complexity of the learner's writing. While it is proficient at identifying simple grammatical issues, it may not catch more nuanced problems like inconsistent voice, awkward phrasing, or register shifts, unless explicitly instructed to do so.

As for transitions, this area often involves logical cohesion across paragraphs, a more global writing skill that can be difficult for AI to assess and explain, especially in narratives where sequencing may not follow typical expository patterns. This observation echoes the findings of Ibrahim and Kirkpatrick (2024), who noted that while ChatGPT is effective at local revisions (e.g., sentence-level edits), it is less adept at global cohesion or genre-specific discourse moves.

These limitations suggest that AI feedback may be most effective when paired with explicit instruction or human guidance on higher-level writing strategies. Teachers might, for instance, use ChatGPT feedback as a springboard for further discussion in writing workshops or peer review sessions.

### ***Cultural and Contextual Limitations of AI Feedback***

Another important finding concerns ChatGPT's lack of cultural awareness and contextual understanding, which some students mentioned as a shortcoming in the qualitative data. For example, when students included cultural references, idiomatic expressions, or regionally specific stories in their narrative essays, ChatGPT occasionally offered feedback that was either irrelevant or overly generic.

This limitation has been noted in earlier research as well (Zebua & Katemba, 2024; Mahapatra, 2024). Because ChatGPT's training data is predominantly drawn from English-speaking contexts, its responses may reflect culturally biased assumptions, and its feedback may not always align with localized communicative norms. This concern is particularly

relevant in narrative writing, where personal experiences, cultural references, and identity expression are central to meaning-making.

To address this issue, scholars have advocated for a hybrid feedback model (Yan, 2024) in which AI-generated feedback is used in conjunction with teacher commentary or peer support. In such models, ChatGPT can take on the role of first-response feedback, while human instructors can supplement with culturally nuanced advice and higher-order rhetorical feedback. This blended approach could maximize the efficiency of AI tools while preserving the pedagogical and interpersonal strengths of human instruction.

### ***Pedagogical Implications***

The results of this study hold several important implications for language educators and curriculum designers. First, the integration of tools like ChatGPT in writing courses can serve as a scalable and cost-effective means of expanding access to feedback, especially in contexts with limited teaching resources. Second, the AI's ability to enhance voice and style, often the most challenging elements to teach in EFL writing, suggests that ChatGPT can play a meaningful role in helping students discover and refine their unique authorial identity.

Educators can also leverage ChatGPT to promote learner autonomy, encouraging students to take initiative in editing their drafts, reflecting on feedback, and setting personalized writing goals. This aligns with contemporary views of language learners as active agents in their own learning process, rather than passive recipients of correction.

However, teacher training will be critical for effective integration. Instructors must be equipped not only to understand how AI tools work but also how to scaffold their use pedagogically, guiding students in interpreting, applying, and evaluating the feedback they receive. Ethical considerations also need to be addressed, including concerns around data privacy, academic integrity, and equitable access to technology.

### ***Limitations and Future Directions***

Although this study provides encouraging evidence for the use of ChatGPT in EFL narrative writing, several limitations must be acknowledged. First, the sample size was relatively small ( $n = 16$ ), and the participants were from a single institution. Future studies should involve larger and more diverse cohorts to increase generalizability. Second, the study focused exclusively on narrative essays; future research could explore other genres such as argumentative, descriptive, or reflective writing to determine whether ChatGPT performs equally well across genres.

Additionally, longitudinal studies are needed to assess sustained improvements in writing quality over time and to explore whether frequent use of AI feedback fosters deeper learning or dependency. Investigating collaborative models, such as combining ChatGPT with peer review or teacher conferencing, could also provide valuable insights into optimal instructional strategies.

## **CONCLUSION**

This study has demonstrated the effectiveness of ChatGPT as a formative feedback tool in enhancing the narrative writing skills of English as a Foreign Language (EFL) learners. By examining pre- and post-feedback writing samples from a group of Indonesian undergraduate students, the findings showed statistically significant improvements in several key aspects of narrative writing, particularly in style, voice, and overall coherence. These results highlight ChatGPT's capacity to go beyond surface-level grammar corrections and support learners in refining the expressive and creative dimensions of their writing, which are essential in the narrative genre.

One of the most notable contributions of this study is its emphasis on narrative writing, a genre that remains underrepresented in existing AI feedback research. Whereas prior studies have largely focused on expository or academic writing, this study adds to the literature by showing that ChatGPT can also be a helpful tool in guiding students to craft personal and emotionally resonant stories. The AI feedback enabled students to reflect more deeply on their tone, language, and storytelling structure, leading to more authentic and cohesive compositions.

In addition to its linguistic impact, ChatGPT was also found to enhance learners' confidence and sense of autonomy. Students reported that the feedback was clear, actionable, and easy to understand, which helped them revise their work independently. This aligns with contemporary language learning frameworks that emphasize the importance of learner agency, self-regulation, and formative assessment. By providing immediate and individualized feedback, ChatGPT empowers learners to take a more active role in the writing and revision process, an especially valuable feature in contexts where teacher feedback may be limited by time or resources.

The study also underscores ChatGPT's potential as a scalable and accessible solution in under-resourced educational environments. In many EFL settings, instructors face large class sizes and cannot provide extensive personalized feedback to each student. ChatGPT helps bridge this gap by offering on-demand support that is not only cost-effective but also consistent in delivery. While it is not a replacement for teacher input, it serves as a valuable complement, particularly in early drafts or low-stakes assignments where rapid feedback can encourage continued engagement with the writing process.

However, the study also acknowledges some limitations of AI-generated feedback, particularly in areas requiring cultural sensitivity and nuanced contextual understanding. Some students found that the feedback occasionally missed the mark when dealing with culturally specific references or subtleties in voice and narrative structure. As such, the integration of ChatGPT into writing instruction should be approached with a hybrid mindset, blending AI support with teacher guidance and peer feedback to ensure comprehensive and contextually appropriate learning experiences.

Moving forward, future research should expand the scope of inquiry by examining ChatGPT's impact across various writing genres, educational levels, and cultural contexts. Longitudinal studies could also investigate the sustainability of writing improvements over time, while experimental designs could compare AI feedback with teacher or peer-based approaches. Ultimately, this study provides encouraging evidence that, when thoughtfully implemented, AI tools like ChatGPT can play a meaningful role in fostering EFL students' writing development in both cognitive and affective domains.

## REFERENCES

- Afifi, S., Rahimi, M., & Wilson, J. (2023). Student engagement with teacher and automated written corrective feedback on L2 writing: a multiple case study. *The Jalt Call Journal*, 19(2), 216-242. <https://doi.org/10.29140/jaltcall.v19n2.1041>
- Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, 21(1), 23. <https://doi.org/10.1186/s41239-024-00455-4>
- Chadha, A. (2024). Transforming higher education for the digital age. *Journal of Interdisciplinary Studies in Education*, 13(S1). <https://doi.org/10.32674/em2qsn46>
- Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Williams, M. D. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges, and implications of generative conversational AI for

- research, practice, and policy. *International Journal of Information Management*, 71, 102642. <https://doi.org/10.1016/j.ijinfomgt.2023.102642>
- Ekanayaka, W.I., & Ellis, R. (2024). Engaging with written corrective feedback. *Instructed Second Language Acquisition*, 8(1), 41-63. <https://doi.org/10.1558/isla.26982>
- Ekizoğlu, M. and Demir, A. (2025). AI-assisted writing feedback for enhancing secondary students' writing skills: an experimental study. *Research Square* <https://doi.org/10.21203/rs.3.rs-6430737/v1>
- Fisher, D. P., Brotto, G., Lim, I., & Southam, C. (2025). The Impact of Timely Formative Feedback on University Student Motivation. *Assessment & Evaluation in Higher Education*, 50(4), 622–631. <https://doi.org/10.1080/02602938.2025.2449891>
- Godwin-Jones, R. (2024). Distributed agency in second language learning and teaching through generative AI. *Language Learning & Technology*, 28(2), 5–30. <https://doi.org/10.1257/73570>
- Hamzah, I. (2023). Students' perception during negotiation of meaning in unfocused and focused tasks. *English Learning Innovation*, 4(1), 1-11. <https://doi.org/10.22219/englie.v4i1.22513>
- Huang, R. (2024). Design and implementation of English writing aids based on natural language processing. *JES: Journal of Electrical Systems*, 20(6s), 2178-2187. <https://doi.org/10.52783/jes.3132>
- Ibrahim, K. and Kirkpatrick, R. (2024). Potentials and implications of ChatGPT for ESL writing instruction. *The International Review of Research in Open and Distributed Learning*, 25(3), 394-409. <https://doi.org/10.19173/irrodl.v25i3.7820>
- Jiang, F., & Hyland, K. (2025). Metadiscursive nouns in academic argument: ChatGPT vs student practices. *Journal of English for Academic Purposes*, 75, 101514. <https://doi.org/10.1016/j.jeap.2025.101514>
- Joe, A. & Mtsi, N. (2024). Impact of teacher workload on academic excellence and learners' metacognition: a case study of two high schools in Chris Hani District. In Tezer, M. (Ed.). *Metacognition in Learning-New Perspectives*. IntechOpen, <https://doi.org/10.5772/intechopen.114316>
- Kara, P., Ognjanović, I., Hölscher, D., Šćekić, L., Kovačević, P., Mantas, J., ... & Bokor, L. (2024). Human-centric digitization in Montenegro: progress through 17 years of national independence and future trends. *Electronics*, 13(13), 2460. <https://doi.org/10.3390/electronics13132460>
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11(1), 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Mao Z, Lee I, Li S.(2024). Written corrective feedback in second language writing: A synthesis of naturalistic classroom studies. *Language Teaching*, 57(4):449-477. <https://doi.org/10.1017/S0261444823000393>
- Ngoi, S., Tan, K., Alias, J., & Mat, N. (2024). Digital storytelling to improve English narrative writing skills. *International Journal of Academic Research in Business and Social Sciences*, 14(4). <https://doi.org/10.6007/ijarbss/v14-i4/21249>
- OpenAI. (2022, November 30). Introducing ChatGPT. *OpenAI*, <https://openai.com/blog/chatgpt>
- Pariyanto, Abdullah, S., Nugroho, M., Barata, F.A., Pramesti, T., Putri, E.P., Ahmad, B., Akrim, and Rudianto (2025). From Classroom to Screen: Examining Academic Achievement in Indonesian Students' Transition to Online Learning in Pre- and Post-



- pandemic. *Pertanika: Journal of Social Science and Humanities*, 33(1), 465 – 481. <https://doi.org/10.47836/pjssh.33.1.21>
- Pariyanto & Tungka, N.F. (2024) Enhancing Writing Skills of EFL Learners Through Automated Feedback: An Empirical Investigation. *Jurnal Ilmiah Spectral*, 10(1), 028-042. <https://doi.org/10.47255/gymtcf78>
- Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. *Cogent Education*, 11(1). <https://doi.org/10.1080/2331186X.2024.2410101>
- Saaty, A. (2023). Correlations between expressing feelings, conveying thoughts, and gaining confidence when writing personal narratives in one's first and second language. *World Journal of English Language*, 13(1), 390. <https://doi.org/10.5430/wjel.v13n1p390>
- Steinert, S., Avila, K.E., Ruzika, S., Kuhn, J., and Küchemann, S. (2024). Harnessing large language models to develop research-based learning assistants for formative feedback. *Smart Learning Environments*, 11(62), 1-15. <https://doi.org/10.1186/s40561-024-00354-1>
- Yan, D. (2024). Comparing individual vs. collaborative processing of ChatGPT-generated feedback: Effects on L2 writing task improvement and learning. *Language Learning & Technology*, 28(1), 1–19. <https://hdl.handle.net/10125/73597>
- Yan, C., He, C., & Sheng, H. (2024). ‘Grades alone are insufficient!’ Chinese EFL student teachers’ perspectives on teacher written feedback on course essays. *Innovations in Education and Teaching International*, 1–15. <https://doi.org/10.1080/14703297.2024.2437116>
- Yunita, W. & Kusuma, L. (2022). Students’ preferences on the focus of feedback in writing research proposals, *Proceedings of the 3rd International Conference on Educational Science and Teacher Profession (ICETeP 2021)* (pp. 141-153). Atlantis Press. [https://doi.org/10.2991/978-2-494069-19-0\\_17](https://doi.org/10.2991/978-2-494069-19-0_17)
- Zebua, J. A. Z., & Katemba, C. V. (2024). Students’ perceptions of using the OpenAI ChatGPT application in improving writing skills. *Journal of Language and Literature Studies*, 4(1), 110–123. <https://doi.org/10.36312/jolls.v4i1.1805>