

The Impact of ChatGPT and Instructor Feedback on ESL Email Writing: A Mixed-Methods Study

*Kesan Maklum Balas ChatGPT dan Pensyarah terhadap Penulisan Emel ESL:
Satu Kajian Kaedah Campuran*

AILEEN FARIDA MOHD ADAM, WAN NAJMIYYAH WAN MD ADNAN &
MOHD SUHAIMI SULAIMAN*

Received: 28-06-2025/Accepted: 18-12-2025

ABSTRACT

This study investigates the impact of ChatGPT-generated feedback and hybrid feedback (ChatGPT + instructor) on ESL students' writing development within a sociocultural and process writing framework. Using a mixed-methods intervention design, 33 Malaysian pre-diploma students completed three email writing tasks: a pre-test (baseline writing), post-test 1 (revised using ChatGPT), and post-test 2 (further revised following instructor feedback). All email responses were assessed using a CEFR-aligned rubric, and students' perceptions were collected via a questionnaire containing both closed- and open-ended items. Results showed moderate gains in writing performance following ChatGPT feedback, particularly in grammar and vocabulary accuracy. However, the most substantial and consistent improvements were observed after hybrid feedback, especially in coherence, content organisation, and task fulfilment. Questionnaire responses revealed students generally appreciated ChatGPT for its immediacy and clarity but still valued instructor feedback for its personalised and contextualised nature. Overall, the findings suggest that combining AI and instructor feedback supports more meaningful revision and writing improvement. This study highlights the pedagogical potential of integrating generative AI into scaffolded writing instruction and recommends future research on hybrid feedback models across broader contexts and proficiency levels.

Keywords: ChatGPT; ESL Writing; Written Feedback; Mixed-Methods; Hybrid Feedback

ABSTRAK

Kajian ini meneroka impak maklum balas yang dijana oleh ChatGPT dan maklum balas hibrid (ChatGPT + pensyarah) terhadap perkembangan penulisan pelajar ESL dalam kerangka sosiobudaya dan penulisan berproses. Menggunakan reka bentuk kajian intervensi kaedah campuran, seramai 33 pelajar pra-diploma dari Malaysia telah melengkapkan tiga tugas penulisan emel: ujian pra (penulisan asas), ujian pasca 1 (disemak semula menggunakan ChatGPT), dan ujian pasca 2 (disemak semula berdasarkan maklum balas pensyarah). Semua esei telah dinilai menggunakan rubrik yang sejajar dengan CEFR, dan persepsi pelajar dikumpulkan melalui soal selidik yang mengandungi item tertutup dan terbuka. Dapatan menunjukkan peningkatan sederhana dalam prestasi penulisan selepas menerima maklum balas ChatGPT, khususnya dalam aspek ketepatan tatabahasa dan perbendaharaan kata. Namun begitu, peningkatan yang paling ketara dan konsisten diperolehi selepas menerima maklum balas hibrid, terutamanya dalam aspek koheren, pengorganisasian kandungan, dan pemenuhan tugas. Responden soal selidik secara amnya menghargai kepantasan dan kejelasan maklum balas daripada ChatGPT, namun tetap menilai tinggi maklum balas pensyarah kerana sifatnya yang bersifat peribadi dan kontekstual. Secara keseluruhannya, kajian ini mencadangkan bahawa gabungan maklum balas AI dan manusia menyokong semakan semula yang lebih bermakna dan peningkatan dalam penulisan. Kajian ini menyerlahkan potensi integrasi AI generatif dalam pengajaran penulisan berasaskan scaffolding dan mencadangkan agar penyelidikan masa depan meneliti model maklum balas hibrid dalam konteks dan tahap kecekapan yang lebih meluas.

Kata kunci: ChatGPT; Penulisan ESL; Maklum Balas Bertulis; Kaedah Campuran; Maklum Balas Hibrid

INTRODUCTION

Written corrective feedback (WCF) has long been central to second language writing pedagogy. WCF refers to teachers' written responses to student errors in areas such as grammar, vocabulary, and organisation, and it has been widely shown to support accuracy and revision in ESL and EFL contexts (Bitchener & Storch, 2016; Hyland & Hyland, 2006). These limitations are particularly evident in ESL classrooms where large class sizes and teacher workload restrict opportunities for personalised feedback.

In response to these challenges, automated written feedback (AWF) has gained prominence as a technological extension of WCF, providing a formative assessment tool that goes beyond summative scoring (Stevenson & Phakiti, 2019). With advancements in natural language processing, AWF systems now include AI-powered tools like Grammarly and ChatGPT, which offer more personalized and interactive feedback (Raheem et al., 2023). While research has explored AWF's effectiveness, user experiences, and impact on writing skills, the findings remain mixed (Shi & Aryadoust, 2024), highlighting the importance of examining how AWF can be integrated with formative feedback strategies to better support student learning (Shi & Aryadoust, 2024).

Emerging evidence suggests that hybrid feedback which combines AI-generated and instructor feedback, may achieve complementary benefits in ESL writing. AI-generated feedback is recognised for its speed, consistency, and effectiveness in improving surface-level aspects of writing such as grammar and vocabulary, while also offering potential to support higher-order concerns including coherence, idea organisation, and content clarity (Bok & Cho, 2023; Shi & Aryadoust, 2024). However, studies also highlight that AI feedback often lacks contextual understanding and task relevance, raising concerns about its reliability in ESL contexts (Najlaa et al., 2024). Zhang et al. (2025) found that while GenAI feedback improved surface-level accuracy, hybrid feedback led to greater improvements in organisation, critical thinking, and overall writing quality among Chinese EFL students. Their findings highlight the value of combining AI with instructor input to achieve more balanced writing support.

In English as a Second Language (ESL) settings such as Malaysia, where English proficiency levels vary widely among tertiary students, formative feedback plays a critical role in writing development (McCarthy, 2022). However, large class sizes and time constraints often limit instructors' capacity to provide timely, individualised feedback (Molin et al., 2020). In this setting, AI tools like ChatGPT offer scalable assistance, particularly for surface-level aspects such as grammar and vocabulary. However, these tools tend to fall short in addressing higher-order concerns like coherence, content organisation, and task fulfilment (Shi & Aryadoust, 2024; Wang et al., 2024). However, little is known about the effectiveness of such hybrid approaches in Southeast Asia, where learners' proficiency levels, digital access, and attitudes toward AI may differ from Western contexts (Nur Ashfaraliana et al., 2024).

This study addresses this knowledge gap by investigating hybrid feedback in a Malaysian ESL context. Specifically, it examines how ChatGPT and instructor feedback, individually and in combination, support the writing development of pre-diploma learners engaged in email writing, an authentic, genre-based task. Grounded in Sociocultural and Process Writing theories, the study also explores learners' perceptions of ChatGPT to provide insights into how AI feedback is received in localised instructional settings. Accordingly, the study was guided by the following objectives:

RO1: To examine the extent to which ChatGPT and instructor feedback, individually and in combination, enhance the writing quality of ESL students.

RO2: To explore ESL students' perceptions of ChatGPT as a writing feedback tool.

To achieve these objectives, the study is guided by the following research questions:

RQ1: To what extent do ChatGPT and instructor feedback, individually and in combination, enhance ESL students' writing quality?

RQ2: What are ESL students' perceptions of ChatGPT as a writing feedback tool?

LITERATURE REVIEW

THE ROLE OF FEEDBACK IN ESL WRITING

Feedback plays a crucial role in the development of writing skills, especially in second language acquisition (SLA), where it helps students improve accuracy, fluency, and overall writing competence (Hyland & Hyland, 2001). Through feedback, learners gain clearer insight into their strengths and areas for improvement, making it a vital component of effective writing instruction.

Despite its importance, providing timely and meaningful formative feedback remains a challenge in many ESL classrooms. Large class sizes, limited contact hours, and teacher workload often reduce the frequency and quality of feedback students receive (Molin et al., 2020; Yao et al., 2024). As a result, many students miss out on the kind of guided support that would help them progress in their writing.

Beyond correction, feedback also supports learner autonomy by encouraging self-regulated learning. When students engage meaningfully with feedback, they become more aware of their writing processes and are better able to revise independently (Zumbrunn et al., 2016). Studies show that effective feedback increases engagement, aids in the internalisation of writing conventions, and leads to measurable improvements over time (Storch & Wigglesworth, 2010; Turner, 2023).

COMPARING AI-GENERATED AND HUMAN INSTRUCTOR FEEDBACK IN L2 WRITING

AI-generated feedback offers immediate, scalable, and consistent support for improving surface-level writing features such as grammar and vocabulary (Raheem et al., 2023; Shadiev & Feng, 2023). Tools like Grammarly and Criterion enhance revision efficiency through rule-based corrections but often fall short when it comes to addressing higher-order concerns such as coherence, task fulfilment, and rhetorical organisation (Bok & Cho, 2023; Shi & Aryadoust, 2024). Generative AI models such as ChatGPT offer more flexible and context-aware suggestions that extend to idea development and text organisation (Mahapatra, 2024). However, their output can still be inconsistent and insufficient for supporting deeper aspects of writing development, particularly in context-sensitive ESL settings (Najlaa et al., 2024).

In contrast, instructor feedback is widely recognised for its personalised nature and contextual sensitivity, which support both linguistic precision and content development (Taskiran & Goksel, 2022). Nonetheless, providing timely and individualised feedback can be challenging in large ESL classrooms where instructors face considerable time and workload constraints (Yao et al., 2024). Moreover, without instructional guidance, students may struggle to apply AI-generated suggestions meaningfully especially when revising meaning-level issues such as coherence or argument structure (Storch & Wigglesworth, 2010).

To bridge these limitations, hybrid feedback models have gained attention as a way to combine the immediacy of AI with the pedagogical depth of instructor input. Zhang et al. (2025), for instance, propose a dialogic, concurrent hybrid approach that blends AI and teacher feedback in real time. In contrast, the present study adopts a process-oriented model where ChatGPT feedback initiates revision and is later deepened through instructor input. Grounded in Sociocultural Theory (Vygotsky, 1978), this sequential scaffolding positions feedback as a mediational tool that supports learners' movement within the Zone of Proximal Development (ZPD). By first engaging with AI-generated feedback and subsequently refining their writing through instructor guidance, students progress from assisted to more independent control of writing processes. This approach balances efficiency with depth, offering a more comprehensive pathway for L2 writing improvement.

CONCEPTUAL FRAMEWORK: HYBRID WRITTEN FEEDBACK MODEL

Building on the principles of Sociocultural Theory (Vygotsky, 1978) and Process Writing Theory (Flower & Hayes, 1981), this study proposes a Hybrid Written Feedback Model to explain how AI-generated and instructor feedback work together to support ESL writing development. The model situates both ChatGPT and instructor input within Vygotsky's ZPD, where ChatGPT provides immediate support for lower-level concerns such as grammar and vocabulary, while instructor feedback extends learning by addressing higher-order elements like coherence, content, and organisation.

In this study, "hybrid feedback" is defined as a sequential process: students first engage with ChatGPT-generated feedback before receiving instructor feedback on the same writing task. Unlike dialogic or concurrent hybrid models (e.g., Zhang et al., 2025), this approach reflects a phased scaffolding strategy embedded within a process writing cycle. The sequencing is informed by both pedagogical reasoning and empirical evidence. For instance, Tran (2025) found that starting with AI feedback allows learners to manage surface-level revisions independently, thereby reducing cognitive load and preparing them for deeper, instructor-guided improvements.

This feedback progression aligns with Process Writing Theory, which views writing as a recursive cycle involving planning, drafting, revising, and editing. By integrating AI at earlier stages and human feedback at later ones, the model supports both the efficiency and depth of revision. Figure 1 illustrates how the model facilitates movement from dependent to independent writing competence through scaffolded feedback across writing phases.

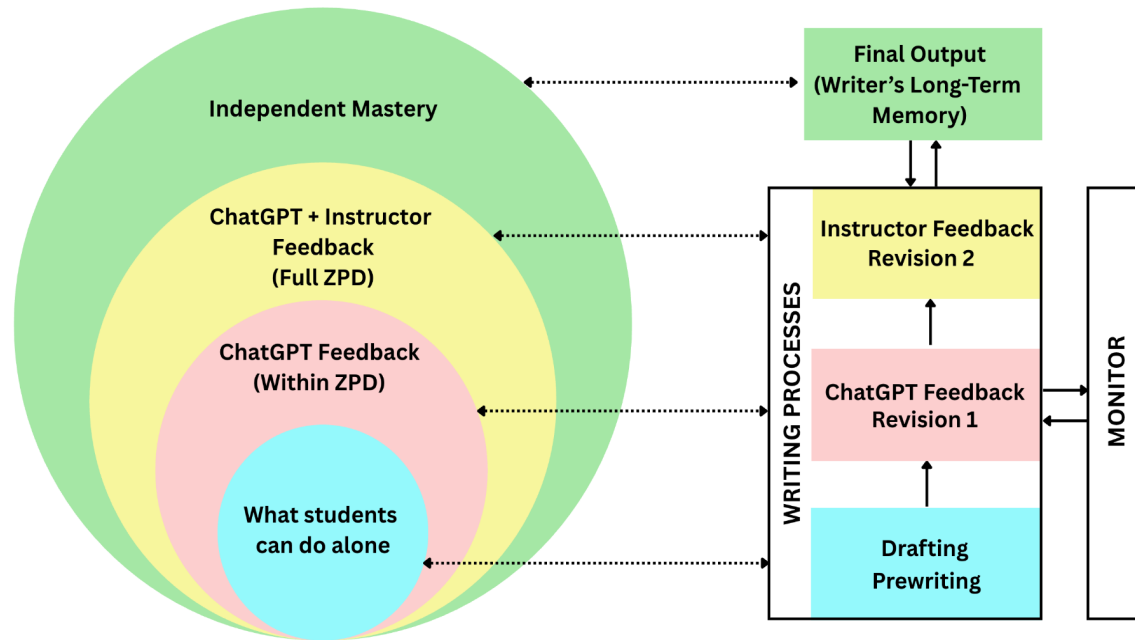


FIGURE 1. Hybrid Feedback Model for ESL Writing Development

The Hybrid Feedback Model integrates Vygotsky's ZPD and Process Writing Theory to illustrate how layered feedback supports writing development. On the left, it shows how students progress from independent writing to scaffolded support: ChatGPT feedback first addresses surface-level issues within the ZPD, followed by instructor feedback that extends revisions to higher-order concerns (full ZPD). On the right, the recursive writing cycle unfolds through drafting, AI-assisted revision, and instructor-guided refinement, with the monitor symbolising students' self-regulation across stages. As learners progress, they move towards greater autonomy and long-term writing competence, with Independent Mastery conceptualised as the ultimate goal of feedback research and a direction for future empirical investigation.

METHODOLOGY

RESEARCH DESIGN

This study employed a mixed-methods intervention design (Creswell & Plano Clark, 2017) to investigate the effectiveness of ChatGPT as an AWF tool in supporting ESL writing development. Guided by the Hybrid Feedback Model, which integrates principles from Vygotsky's Sociocultural Theory and Process Writing Theory, the intervention was structured around three recursive phases: initial drafting, AI-assisted revision, and instructor-guided refinement.

The choice of a mixed-methods design is justified by the dual focus of the study. First, the quantitative strand examined measurable changes in writing quality across different feedback conditions, thereby addressing the objective of determining the extent to which ChatGPT and instructor feedback improved writing performance. Second, the qualitative strand explored learners' perceptions and experiences, providing contextual insights into how feedback was interpreted and applied in practice. This integration of statistical trends with learner voices allowed

for a more comprehensive understanding of both the outcomes and processes of writing development, consistent with recommendations for educational intervention research (Gopalan et al., 2020).

Although the two cohorts were enrolled in different semesters, they shared comparable academic backgrounds, course objectives, and English language proficiency levels. As such, their writing performance data were combined for analysis related to Research Question 1 (RQ1). For Research Question 2 (RQ2), which explored students' perceptions of ChatGPT feedback, only data from the first cohort ($n = 13$) were analysed. In perception-based research, this approach is considered as methodologically acceptable, especially when data collection was carried out under consistent conditions and the objective is to produce exploratory insights rather than to generalise the findings (Creswell & Clark, 2017).

Thirty-three Malaysian pre-diploma students completed three writing tasks across the study. Prior to using ChatGPT, students received structured training, supporting research that highlights the importance of preparatory instruction when integrating digital tools (Mathur & Mahapatra, 2022). This also aligns with Widad et al. (2024), who emphasise the pedagogical role of digital competence in higher education.

PARTICIPANTS

This study involved 33 pre-diploma Business English students from a Malaysian public university, recruited through purposive sampling. Business English students were chosen because email writing is a core component of their curriculum, making them an appropriate population for examining authentic, genre-based writing tasks. Two consecutive cohorts were included: March–August 2024 ($n = 13$) and October 2024–February 2025 ($n = 21$), although one student in the latter cohort did not consent, resulting in 20 participants.

Inclusion criteria required that participants (i) were enrolled in the Business English course, (ii) had completed the pre-test writing task, and (iii) consented to participate in the study. Based on their pre-test performance, students ranged from CEFR A1 to B2 levels, indicating overall low to intermediate proficiency. All students completed identical email writing tasks, received both ChatGPT and instructor feedback, and were assessed using a common rubric by trained raters. Participation was voluntary, and students were assured of confidentiality and the right to withdraw at any time.

Although the two cohorts were drawn from different semesters, they shared comparable academic backgrounds, course objectives, and proficiency profiles. Their writing performance data were therefore combined for Research Question 1 (RQ1). For Research Question 2 (RQ2), which examined perceptions of ChatGPT, only the first cohort ($n = 13$) was included. In exploratory perception studies, analysing a single cohort under consistent conditions is considered methodologically acceptable (Creswell & Clark, 2017).

RESEARCH INSTRUMENTS

A total of 99 email responses were collected and analysed across three test phases to examine the impact of ChatGPT and hybrid feedback on ESL writing performance (RQ1). Each student produced three drafts: a Pre-Test (baseline draft), a Post-Test 1 email (revised using ChatGPT), and a Post-Test 2 email (further revised following instructor feedback). All responses were assessed with a CEFR-aligned rubric comprising four components: task fulfilment, language

accuracy, vocabulary range, and coherence, allowing analysis of both surface-level and higher-order writing skills.

To address RQ2, a self-developed questionnaire was administered to explore students' perceptions of ChatGPT as a feedback tool. The instrument originally consisted of 19 items, but analysis in this study focused on nine that were directly aligned with the research question: six close-ended items and three open-ended prompts. The close-ended items measured both students' prior experience with ChatGPT and their perceptions of its usefulness for writing support. Specifically, they examined the frequency of use before class, self-rated ability to use ChatGPT for writing, ChatGPT's role in helping students write essays and articles, its value as a tool for productivity, its contribution to generating ideas for writing, and its importance in helping students find answers in a university context. These items were rated on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The three open-ended items complemented these measures by inviting students to elaborate on the benefits of receiving paragraph writing feedback through ChatGPT, the problems encountered when using ChatGPT feedback, and their preference for receiving feedback from a lecturer face-to-face versus from an AI application such as ChatGPT.

To enhance content and face validity, the questionnaire was reviewed by two TESL colleagues with expertise in language pedagogy and educational technology, and minor revisions were made for clarity and contextual appropriateness. To evaluate internal reliability, Cronbach's alpha was calculated for the six Likert-scale items, yielding a coefficient of 0.75, which indicates acceptable internal consistency (Tavakol & Dennick, 2011).

In addition, the custom GPT used in this study, *Email Writing Buddy*, was designed via OpenAI's platform to align with the assessment rubric. Figure 2 displays the landing page interface presented to students. The GPT's behaviour was defined through customized instructions specifying its role, feedback strategy, and alignment with the course rubric. The knowledge section included embedded images of the writing prompt and rubric to ensure contextual accuracy. To streamline interaction, the study adopted OpenAI's default conversation starters, which intuitively guided students through the feedback process.

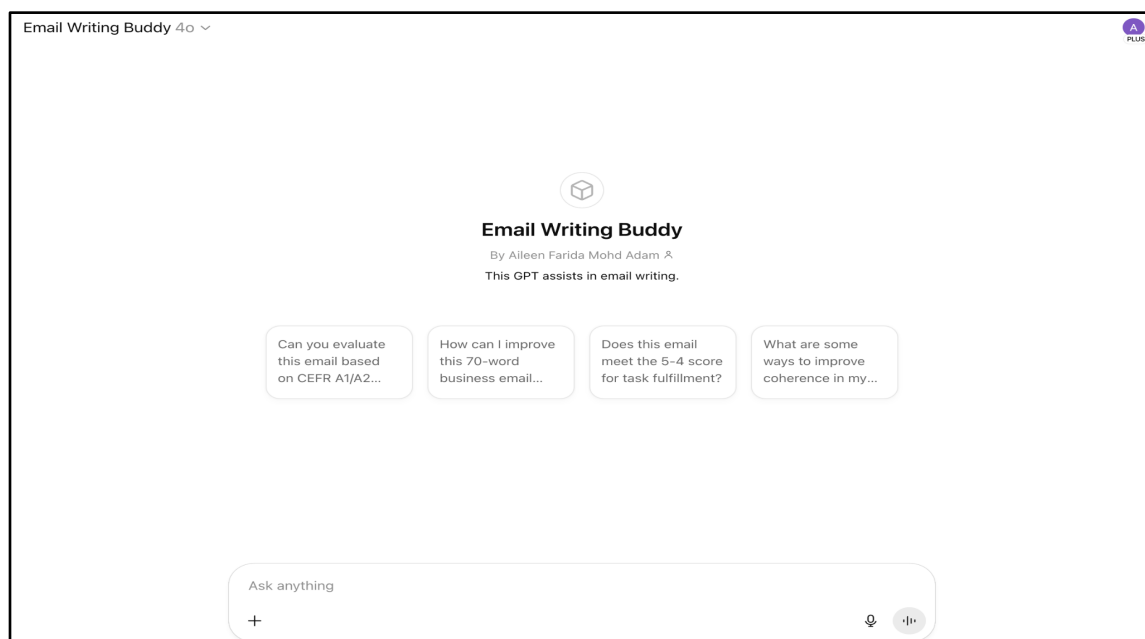


FIGURE 2. The landing page of the custom GPT 'Email Writing Buddy'

To guide students' second revision, instructor feedback was provided following their initial ChatGPT-assisted drafts, as shown in Figure 3.

Suraya received an email from Jamil, asking for the details of the phone that Suraya is selling. Help Suraya reply to the email.

- Include the following points:
- the price of the phone ✓
 - the phone's model ✗
 - the phone's storage (in GB) ✗
 - the colour of the phone ✓
 - the time and place to meet up ✗
- Handwritten feedback:*
- You need to write 70-80 words
 - Include words that connect ideas
such as: and, but, because, so...
 - so your ideas will become clearer
 - don't forget to use polite language
to make you sound more approachable
 - Also include all required points

FIGURE 3. Example of instructor feedback

Instructor's written feedback were provided after students had revised their drafts using ChatGPT. As students were free to use Email Writing Buddy in various ways without specific prompt restrictions, the instructor feedback focused on areas still lacking or underdeveloped in the revised drafts. This approach ensured that feedback addressed residual weaknesses, offering support where automated feedback may have been insufficient.

PROCEDURE AND DATA ANALYSIS

The instructional and assessment sequence (see Table 1) was designed to reflect the recursive nature of process writing and the scaffolded feedback progression outlined in the Hybrid Feedback Model. Each student completed three writing tasks: a Pre-Test, a ChatGPT-assisted revision (Post-Test 1), and a final revision with instructor feedback (Post-Test 2). Given the students' low to intermediate proficiency levels (A1–B2), the hybrid feedback approach was intended to provide scaffolded support suited to varying learner needs. Writing performance data informed Research Objective 1 (RO1), while student perceptions were collected through a survey to address Research Objective 2 (RO2).

PRE-TEST (WEEK 2)

Students completed a handwritten email task in class within a one-hour time limit. The prompt required them to respond to an email from an interested buyer requesting information about a phone for sale, addressing five specific content points (e.g., price, model, colour, etc.). No rubric guidance or writing instruction was provided beforehand, ensuring the Pre-Test served as a diagnostic assessment of students' independent writing ability. The email responses were later graded using a CEFR-aligned rubric to determine each student's proficiency level, which ranged from A1 to B2. According to Weigle (2002), initial writing tasks conducted without scaffolding or intervention provide a useful baseline to assess learners' unaided writing proficiency and identify instructional needs. Following this, students received four hours of writing instruction and two hours of ChatGPT training before beginning the revision process.

POST-TEST 1 (WEEK 5 – CHATGPT FEEDBACK)

Students revised their original Pre-Test email drafts and also produced new versions based on feedback from *Email Writing Buddy*, a custom GPT aligned with the CEFR-based rubric. They accessed the tool via their mobile phones using a shared browser link. While instructor-designed sample prompts were provided, students were encouraged to interact with ChatGPT autonomously, asking follow-up questions to clarify or extend the feedback received. Promoting digital competence and learner agency through guided use of AI tools enhances students' capacity to engage meaningfully in technology-mediated tasks (Widad et al., 2024). Handwritten drafts were uploaded using ChatGPT's in-app camera or photo upload feature. All revisions were completed in class, and both original and revised versions were submitted to the instructor. The instructor provided technical assistance and clarified rubric-related queries but did not give content-specific feedback during this phase to preserve the integrity of AI-mediated revision.

POST-TEST 2 (WEEK 7 – INSTRUCTOR FEEDBACK)

Two weeks after Post-Test 1, the instructor returned students' ChatGPT-revised drafts, now marked using the CEFR-aligned rubric. Feedback was provided in written form, highlighting missing content, language issues, and rhetorical weaknesses, with comments addressing both surface-level features and broader writing concerns (e.g., use of connectors, tone, and task fulfilment). This stage is best characterised as direct, rubric-aligned, focused feedback rather than metalinguistic correction.

For low proficiency ESL learners, direct, rubric-aligned feedback is often more effective than metalinguistic correction because it reduces cognitive load and increases clarity. Learners at this level may not yet have the metalinguistic awareness or grammatical vocabulary needed to interpret abstract error labels (Ellis, 2009). Therefore, providing explicit comments such as pointing out missing information, suggesting more polite expressions, or highlighting unclear phrasing helps them understand exactly what needs to be improved without requiring advanced language analysis skills.

Students revised their emails in class with the instructor present to clarify feedback and support application, reinforcing the full ZPD layer of the Hybrid Feedback Model. The final revised email responses were collected as Post-Test 2 data.

STUDENT PERCEPTIONS SURVEY (WEEK 7)

Immediately following Post-Test 2, students completed a questionnaire on their experience using ChatGPT for writing feedback. The instrument included six close-ended Likert-scale items and three open-ended questions. Quantitative responses were analysed using frequency counts, while qualitative responses (including those written in Malay) were translated and thematically coded to uncover patterns in perception. Although the writing task focused on email writing, the questionnaire used the broader term “*essay*” to refer to students' writing. This decision reflects the instructional practice at the pre-diploma level where email writing is treated as a structured form of academic writing. For the purpose of this study, students' responses to “*essay*” were interpreted in the context of their recent experience with the email writing task.

A summary of the data collection activities and their alignment with the research objectives is presented in Table 1.

TABLE 1. Data Collection Phases and Linked Objectives

Phase	Week(s)	Key Activities	Linked Research Objective(s)
1	2	Pre-Test - Students wrote an email response (70-80 words) - Email responses graded as diagnostic test	Objective 1 ¹
2	3 - 4	Intervention Email Writing Instruction (4 hours): - Key writing strategies for business-related emails were taught - Rubric criteria: task fulfilment, language accuracy, vocabulary range, coherence, and cohesion was explained ChatGPT Training (2 hours): <i>Email Writing Buddy</i>, a custom GPT aligned with the rubric, was introduced - Students were trained on using prompts and applying revisions effectively.	Objective 1
3	5	Post-test 1 (ChatGPT Feedback): Students submitted pre-test email responses to <i>Email Writing Buddy</i> for rubric-aligned feedback and revised their responses.	Objective 1
4	7	Post-test 2 (ChatGPT + Instructor Feedback): Two weeks after Post-Test 1, the instructor provided written feedback, which students used to revise their email drafts.	Objective 1
		Students' Perceptions Survey Immediately after Post-Test 2, students were asked to answer a questionnaire on ChatGPT feedback experience (quantitative + qualitative questions)	Objective 2 ²

DATA ANALYSIS

For RO1, students' writing scores across the three test phases were analysed using Statistical Package for the Social Sciences (SPSS). Descriptive statistics, paired-sample t-tests, and post hoc comparisons were conducted to examine score trends and statistically significant differences. Inter-rater reliability was assessed using Intraclass Correlation Coefficients (ICC) to ensure scoring consistency between the two trained raters.

For RO2, student perception data were explored through descriptive statistics and inductive thematic analysis following Brown and Clarke's (2006) framework. The analysis began with familiarisation of data through repeated readings of data. Then, initial coding was done to identify meaningful categories related to learners' experiences with AI-generated and instructor's feedback. Next, the codes were used to search for preliminary themes. After that, the themes were reviewed and refined, and the key extracts were selected and reported with supporting quotations.

¹ Objective 1: To investigate the effectiveness of ChatGPT and instructor feedback in enhancing ESL students' writing quality.

² Objective 2: To explore ESL students' perceptions of ChatGPT as a writing feedback tool.

INTER-RATER RELIABILITY

Inter-rater reliability was assessed using the ICC in SPSS to evaluate the consistency between two independent raters, both with 14 years of English teaching experience. Each rater evaluated all 99 email responses across the Pre-Test, Post-Test 1, and Post-Test 2 phases. As shown in Table 2, the ICC values indicate excellent agreement for the Pre-Test (0.947) and Post-Test 1 (0.919), and moderate agreement for Post-Test 2 (0.674). These results confirm generally high consistency in scoring across phases, with some variation during the final rating stage.

TABLE 2. Intraclass Correlation Coefficient Table

Test	Intraclass Correlation Coefficient
Pre-Test	0.947
Post-Test 1	0.919
Post-Test 2	0.674

RESULTS AND DISCUSSION

The findings are organized by research objectives, with writing performance analyzed using SPSS (descriptive statistics, paired t-tests, post hoc tests) and questionnaire responses examined through frequency and thematic coding.

RO1: Comparison of Writing Improvements between ESL Students Receiving ChatGPT Feedback and Those Receiving Combined ChatGPT and Instructor Feedback

TESTS OF NORMALITY

The Tests of Normality (Table 3) show that all three writing phases meet the assumption of normality, as indicated by non-significant Shapiro-Wilk values ($p > .05$). While the Kolmogorov-Smirnov test returned a marginal result for Post-Test 2 ($p = .037$), the Shapiro-Wilk test, which is more appropriate for small sample sizes ($n < 50$), confirms normal distribution across all datasets. These findings support the use of parametric tests in subsequent analyses. Table 3 presents the detailed results.

TABLE 3. Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
<i>Pre_Test_Students_Original</i>	.131	33	.164	.969	33	.462
<i>Post_Test1_ChatGPT_Feedback</i>	.152	33	.051	.952	33	.151
<i>Post_Test2_ChatGPT_Instructor_Feedback</i>	.157	33	.037	.950	33	.136

a. Lilliefors Significance Correction

OVERVIEW OF WRITING PERFORMANCE TRENDS

Students' writing progress across test phases was analyzed using descriptive statistics to examine score variability and performance trends as shown in Table 4.

TABLE 4. Overall Writing Performance across Tests

	<i>N</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Mean</i>	<i>Std. Deviation</i>
<i>Pre_Test_Students_Original</i>	33	8.50	15.50	11.9848	2.31352
<i>Post_Test1_ChatGPT_Feedback</i>	33	10.50	18.00	14.0455	2.13976
<i>Post_Test2_ChatGPT_Instructor_Feedback</i>	33	14.00	19.50	16.6970	.88237
<i>Valid N (listwise)</i>	99	8.50	19.50	14.2424	2.56980

Students' writing performance improved progressively across all test phases, with mean scores rising from 11.98 in the Pre-Test to 14.05 in Post-Test 1 and 16.70 in Post-Test 2. Score variability, as reflected in the range, was highest in the Pre-Test (8.50–15.50), slightly narrower in Post-Test 1 (10.50–18.00), and lowest in Post-Test 2 (14.00–19.50), indicating increased consistency in student performance. The most substantial gain occurred between Post-Test 1 and Post-Test 2, suggesting that combining ChatGPT with instructor feedback led to more meaningful improvements in writing quality.

SCORE DISTRIBUTION AND VARIABILITY

The boxplot (Figure 4) visually illustrates score distribution and variability across the three test phases, highlighting performance trends and consistency.

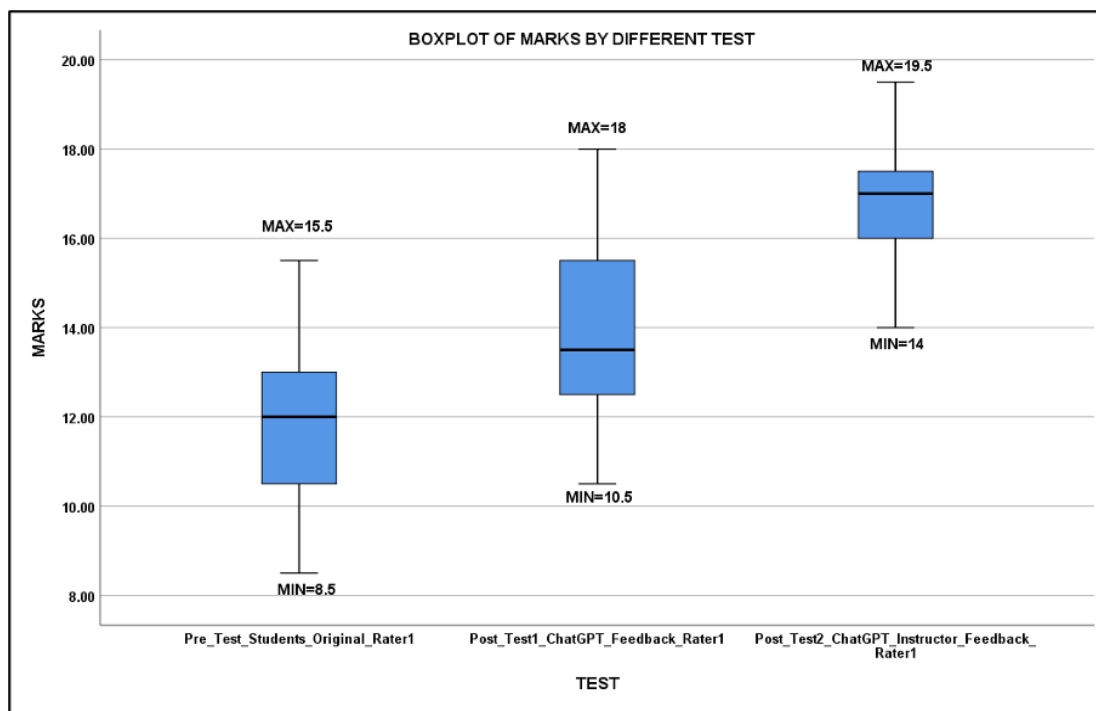


FIGURE 4. Score distributions and variability across the Pre-Test, Post-Test 1, and Post-Test 2

Students' writing performance improved steadily across test phases, with higher median scores and reduced variability. The Pre-Test showed the widest score range (8.5–15.5), indicating diverse initial abilities. Post-Test 1 (ChatGPT feedback) improved scores (11.5–18.5) but retained some variability. Post-Test 2 (ChatGPT + instructor feedback) saw the greatest gains (16.0–19.5), suggesting hybrid feedback led to more consistent and significant improvements, with an outlier indicating strong effectiveness for some students.

OVERALL TRENDS IN WRITING DEVELOPMENT ACROSS PHASES

The Tukey HSD post hoc test (Table 5) confirmed statistically significant improvements ($p = 0.000$) across all test phases.

TABLE 5. Tukey HSD Post Hoc Test

Dependent Variable: MARKS						
Tukey HSD						
(I) TEST	(J) TEST	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Pre_Test_Students_Original_Rater1	Post_Test1_ChatGPT_Feedback_Rater1	-2.06061*	.41961	.000	-3.0595	-1.0617
	Post_Test2_ChatGPT_Instructor_Feedback_Rater1	-4.71212*	.41961	.000	-5.7110	-3.7132
Post_Test1_ChatGPT_Feedback_Rater1	Pre_Test_Students_Original_Rater1	2.06061*	.41961	.000	1.0617	3.0595
	Post_Test2_ChatGPT_Instructor_Feedback_Rater1	-2.65152*	.41961	.000	-3.6504	-1.6526
Post_Test2_ChatGPT_Instructor_Feedback_Rater1	Pre_Test_Students_Original_Rater1	4.71212*	.41961	.000	3.7132	5.7110
	Post_Test1_ChatGPT_Feedback_Rater1	2.65152*	.41961	.000	1.6526	3.6504

*. The mean difference is significant at the 0.05 level.

The greatest improvement in writing scores was observed between the Pre-Test and Post-Test 2, with a statistically significant mean increase of 4.71 points. A 2.06-point gain from Pre-Test to Post-Test 1 reflects the impact of ChatGPT feedback alone, while an additional 2.65-point increase from Post-Test 1 to Post-Test 2 highlights the added value of instructor input. These results underscore the effectiveness of hybrid feedback in enhancing overall writing quality.

The progressive improvement across all writing phases—from initial AI-assisted revision to subsequent hybrid feedback—demonstrates how scaffolded feedback sequencing supports continuous learning. The early gains following ChatGPT feedback reflect learners' responsiveness to immediate, form-focused correction, whereas the more substantial leap after instructor mediation highlights the depth added through dialogic feedback. This developmental pattern parallels the scaffolding process described in Sociocultural Theory, where support gradually shifts from automated assistance to expert guidance. Similar staged progressions were reported in Zhai and Ma (2022), who found that feedback delivered in iterative cycles fosters cumulative learning rather than isolated corrections.

COMPARISON OF WRITING PERFORMANCE BY RUBRIC ELEMENTS

To better understand which aspects of writing improved most and how students responded to different feedback types, individual rubric elements were analyzed. The rubric elements are Task Fulfilment, Grammar and Vocabulary Accuracy, Vocabulary Range, and Coherence and Cohesion. Each subcomponent was examined using descriptive trends (means plots) and statistical comparisons (Tukey HSD) to determine the effectiveness of ChatGPT-only feedback versus combined ChatGPT and instructor feedback.

The following sections present the findings for each element, offering insight into how different aspects of writing were shaped across the feedback stages.

TASK FULFILLMENT

Task Fulfilment refers to the extent to which students addressed the requirements of the email prompt. As shown in the means plot (Figure 5), scores increased progressively across phases, with the most pronounced gain between Post-Test 1 and Post-Test 2. This suggests that while students benefited modestly from ChatGPT feedback, the addition of instructor guidance helped them better align their writing with task expectations.

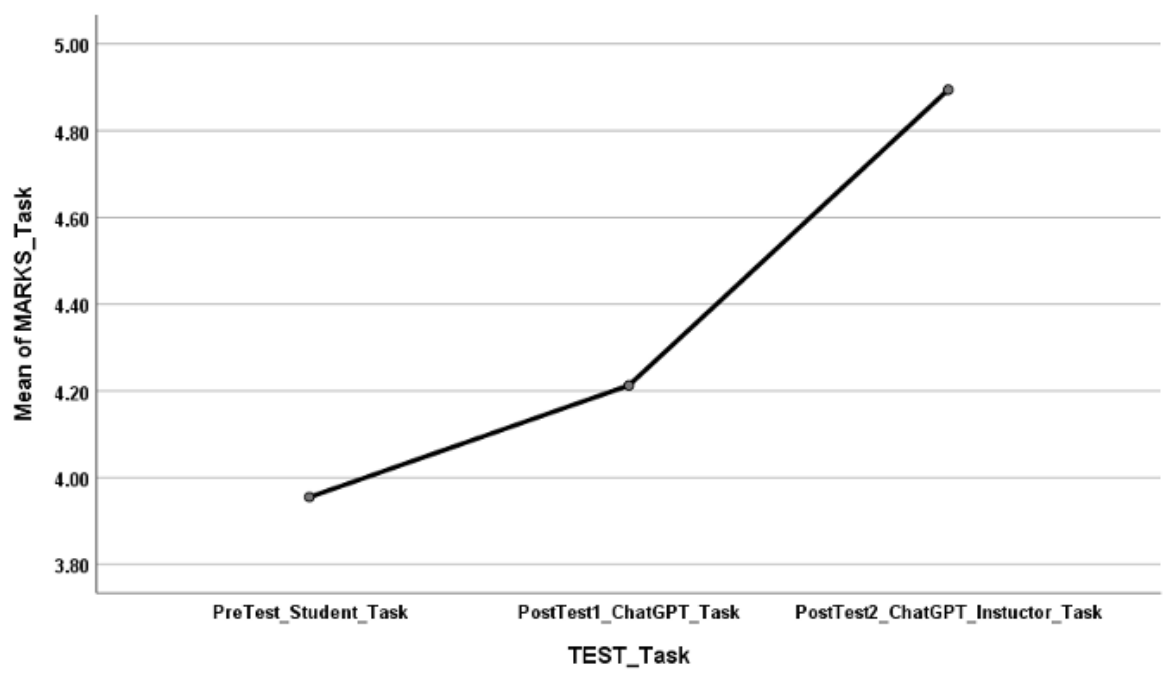


FIGURE 5. Mean Plot of Task Fulfilment Scores Across Test Phases

Tukey HSD results (Table 6) support this trend. The difference between Pre-Test and Post-Test 1 was not statistically significant ($p = .503$), indicating limited effect from ChatGPT alone. However, significant gains were recorded from Pre-Test to Post-Test 2 (mean difference = 0.93, $p < .001$) and from Post-Test 1 to Post-Test 2 (mean difference = 0.68, $p = .010$). This underscores the importance of instructor feedback in clarifying task purpose and content scope.

TABLE 6. Tukey HSD Post Hoc Comparisons for Task Fulfilment Scores Across Test Phases

Multiple Comparisons Dependent Variable: MARKS_Task Tukey HSD						
(I) TEST_Task	(J) TEST_Task	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
PreTest_Student_Task	PostTest1_ChatGPT_Task	-.25758	.22970	.503	-.8044	.2893
	PostTest2_ChatGPT_Instructor_Task	-.93939*	.22970	.000	-1.4862	-.3926
PostTest1_ChatGPT_Task	PreTest_Student_Task	.25758	.22970	.503	-.2893	.8044
	PostTest2_ChatGPT_Instructor_Task	-.68182*	.22970	.010	-1.2286	-.1350
PostTest2_ChatGPT_Instructor_Task	PreTest_Student_Task	.93939*	.22970	.000	.3926	1.4862
	PostTest1_ChatGPT_Task	.68182*	.22970	.010	.1350	1.2286

*. The mean difference is significant at the 0.05 level.

The modest gains in task fulfilment after ChatGPT-only feedback suggest that AI feedback was limited in helping students interpret communicative intent or task purpose. As Shi and Aryadoust (2024) observed, AI-based AWF systems provide effective linguistic scaffolding but lack genre sensitivity and awareness of discourse goals. This likely explains why substantial improvement only occurred after instructor feedback clarified task expectations and rhetorical intent.

GRAMMAR AND VOCABULARY ACCURACY

Grammar and vocabulary accuracy assesses sentence-level correctness and language control. As depicted in the means plot (Figure 6), student scores rose steadily from the initial draft to the final version, with the greatest improvement evident in the hybrid feedback condition.

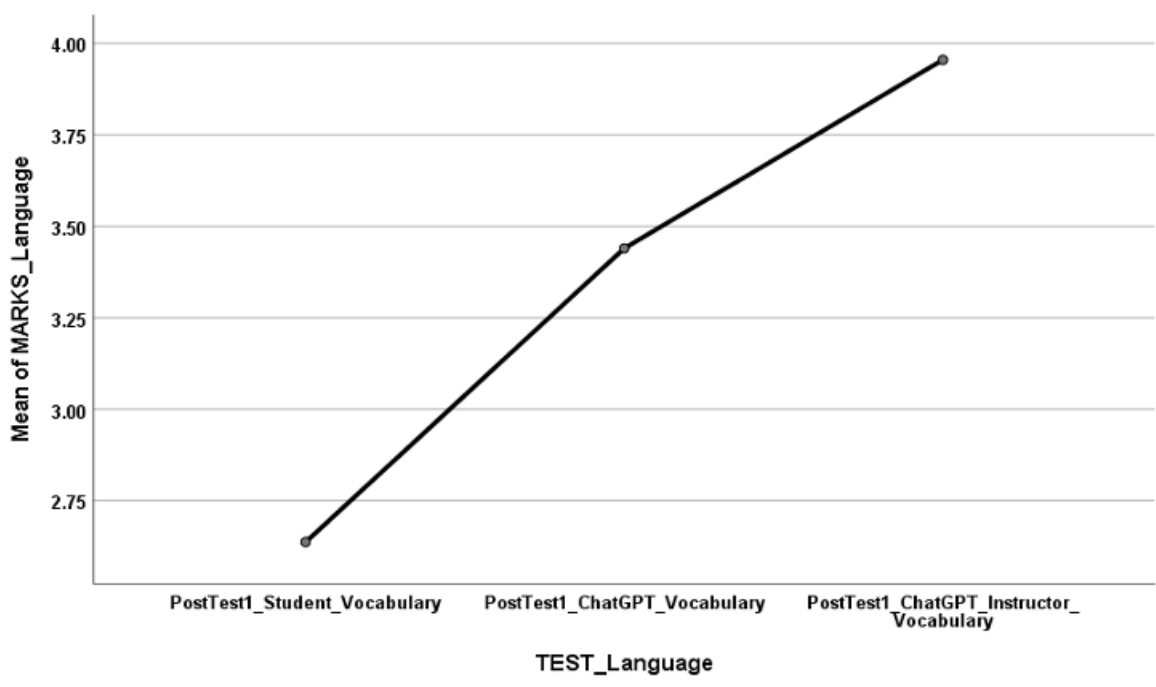


FIGURE 6. Means Plot of Grammar and Vocabulary Accuracy Across Test Conditions

The Tukey HSD test (Table 7) revealed significant differences between all conditions ($p < .001$). The largest gain occurred between the Pre-Test and Post-Test 2 (mean difference = 1.32), followed by improvements from Pre-Test to Post-Test 1 (0.80), and Post-Test 1 to Post-Test 2 (0.52). These results demonstrate that while ChatGPT feedback helped reduce basic grammatical issues, instructor input further refined students' grammatical accuracy and lexical appropriateness.

TABLE 7. Tukey HSD Post Hoc Comparisons of Grammar and Vocabulary Accuracy across Test Conditions

Dependent Variable: MARKS Language						
Tukey HSD						
(I) TEST_Language	(J) TEST_Language	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
PostTest1_Student_Vocabulary	PostTest1_ChatGPT_Vocabulary	-.80303*	.11009	.000	-1.0651	-.5410
	PostTest1_ChatGPT_Instructor_Vocabulary	-1.31818*	.11009	.000	-1.5803	-1.0561
PostTest1_ChatGPT_Vocabulary	PostTest1_Student_Vocabulary	.80303*	.11009	.000	.5410	1.0651
	PostTest1_ChatGPT_Instructor_Vocabulary	-.51515*	.11009	.000	-.7772	-.2531
PostTest1_ChatGPT_Instructor_Vocabulary	PostTest1_Student_Vocabulary	1.31818*	.11009	.000	1.0561	1.5803
	PostTest1_ChatGPT_Vocabulary	.51515*	.11009	.000	.2531	.7772

*. The mean difference is significant at the 0.05 level.

The pronounced reduction in grammatical and lexical errors following ChatGPT feedback confirms that AI systems are particularly effective for surface-level accuracy. Consistent with Zhai and Ma (2022), AWE tools yield large gains in error detection and correction through consistent, form-focused input. Yet, these systems contribute less to the *why* behind language choices. The smaller but meaningful improvement after instructor feedback suggests that teachers help internalise grammatical awareness and lexical appropriateness by contextualising feedback and explaining underlying rules—an observation echoed in Shi and Aryadoust (2024), who noted that human mediation transforms mechanical correction into conceptual understanding.

VOCABULARY RANGE

Vocabulary Range evaluates students' ability to use a diverse and appropriate range of words. As shown in the means plot (Figure 7), students' scores increased across all phases, with a consistent upward trend from Pre-Test to Post-Test 2. This reflects gradual but meaningful growth in lexical sophistication.

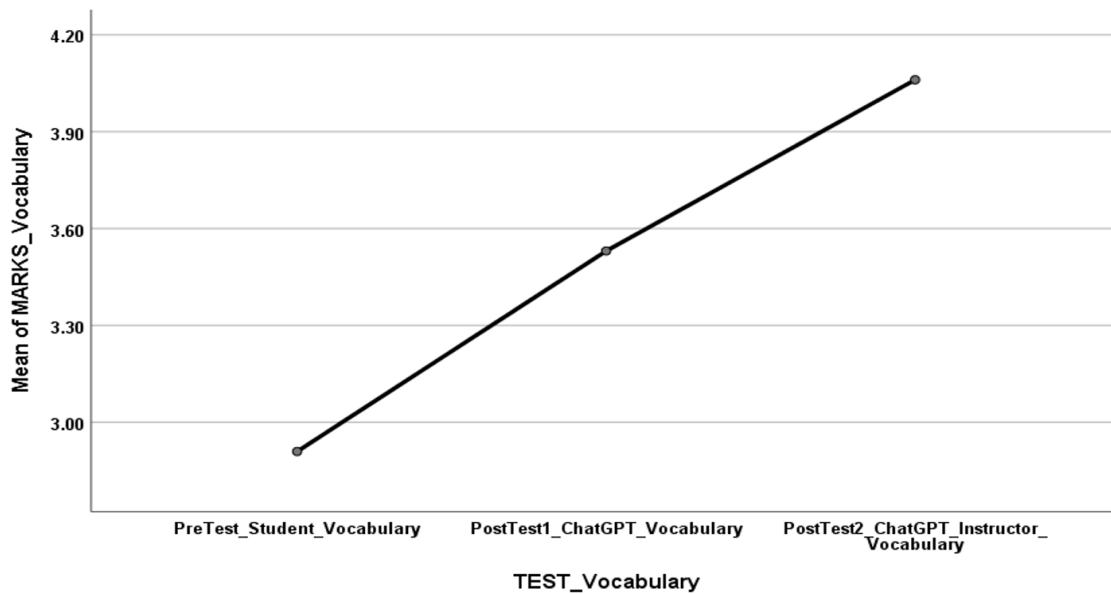


FIGURE 7. Means Plot of Vocabulary Range across Test Phases

Tukey HSD results (Table 8) confirmed that all pairwise differences were statistically significant ($p < .001$). The largest gain was recorded between the Pre-Test and Post-Test 2 (mean difference = 1.15), followed by significant improvements from Pre-Test to Post-Test 1 (0.62) and Post-Test 1 to Post-Test 2 (0.53). These findings suggest that vocabulary development was enhanced most when students received both AI and instructor feedback.

TABLE 8. Tukey HSD results of Vocabulary Range across Test Phases

Dependent Variable: MARKS_Vocabulary						
Tukey HSD						
(I) TEST_Vocabulary	(J) TEST_Vocabulary	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
PreTest_Student_Vocabulary	PostTest1_ChatGPT_Vocabulary	-.62121*	.12329	.000	-.9147	-.3277
	PostTest2_ChatGPT_Instructor Vocabulary	-1.15152*	.12329	.000	-1.4450	-.8580
PostTest1_ChatGPT_Vocabulary	PreTest_Student_Vocabulary	.62121*	.12329	.000	.3277	.9147
	PostTest2_ChatGPT_Instructor Vocabulary	-.53030*	.12329	.000	-.8238	-.2368
PostTest2_ChatGPT_Instructor Vocabulary	PreTest_Student_Vocabulary	1.15152*	.12329	.000	.8580	1.4450
	PostTest1_ChatGPT_Vocabulary	.53030*	.12329	.000	.2368	.8238

*. The mean difference is significant at the 0.05 level.

The gradual increase in vocabulary range across the three writing phases also reflects learners' expanding linguistic exposure through interaction with AI-generated models of language use. ChatGPT appears to function as a lexical scaffold, offering learners varied expressions and synonyms that enrich lexical diversity, an outcome consistent with Najlaa et al. (2024), who found that automated grammar and paraphrasing tools such as Grammarly and QuillBot can enhance vocabulary use and overall text variety in ESL writing.

COHERENCE AND COHESION

Coherence and cohesion reflect how clearly students connect and organize ideas throughout the text. As presented in the means plot (Figure 8), scores improved steadily, with the greatest gain observed after hybrid feedback. This suggests that combining ChatGPT with instructor input strengthened students' ability to structure their writing logically and fluidly.

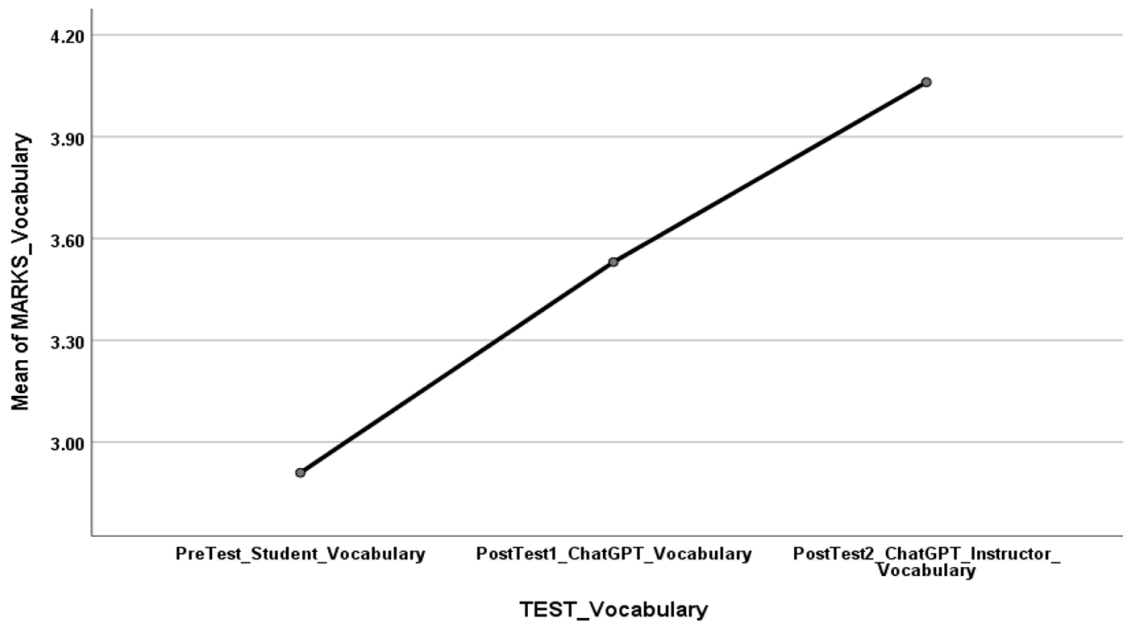


FIGURE 8. Mean Plot of Coherence and Cohesion Scores Across Test Phases

The Tukey HSD analysis (Table 9) showed statistically significant differences across all test phases ($p < .001$). The largest improvement was from Pre-Test to Post-Test 2 (mean difference = 1.15), followed by gains between Pre-Test and Post-Test 1 (0.62) and between Post-Test 1 and Post-Test 2 (0.53). These results mirror those of vocabulary and language accuracy, reinforcing the enhanced value of instructor-led scaffolding.

TABLE 9. Tukey HSD Post Hoc Comparisons for Coherence and Cohesion Scores

Dependent Variable: MARKS Coherence						
Tukey HSD						
(I) TEST_Coherence	(J) TEST_Coherence	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
PreTest_Student_Coherence	PostTest1_ChatGPT_Coherence	-.37879	.20318	.155	-.8625	.1049
	PostTest2_ChatGPT_Instructor Coherence	-1.30303*	.20318	.000	-1.7867	-.8193
PostTest1_ChatGPT_Coherence	PreTest_Student_Coherence	.37879	.20318	.155	-.1049	.8625
	PostTest2_ChatGPT_Instructor Coherence	-.92424*	.20318	.000	-1.4079	-.4405
PostTest2_ChatGPT_Instructor Coherence	PreTest_Student_Coherence	1.30303*	.20318	.000	.8193	1.7867
	PostTest1_ChatGPT_Coherence	.92424*	.20318	.000	.4405	1.4079

*. The mean difference is significant at the 0.05 level.

The sharp improvement in coherence and cohesion following hybrid feedback underscores the shift from local correction to global text management. Whereas AI feedback appeared to assist primarily with surface connectors and sentence-level transitions, instructor feedback enabled learners to reorganise ideas and achieve logical flow throughout the text. This pattern mirrors the collaborative scaffolding mechanism described in Sociocultural Theory, where guidance from a more knowledgeable other supports learners in mastering complex, global writing skills. Empirical evidence from Zhang et al. (2025) likewise showed that hybrid feedback combining AI and teacher input led to greater improvements in L2 writing quality than AI feedback alone. In our data, the hybrid stage therefore represents not just an additive effect but a qualitative leap: from editing isolated sentences to shaping cohesive discourse.

SUMMARY OF FINDINGS AND ANALYSIS FOR RESEARCH OBJECTIVE 1

Across all four rubric elements, students' writing performance improved progressively from the Pre-Test to Post-Test 2, with the greatest gains observed following hybrid feedback. While ChatGPT feedback alone facilitated moderate improvements, particularly in Grammar and Vocabulary Accuracy, it was the combination of ChatGPT and instructor feedback that resulted in the most substantial and consistent enhancements (Mahapatra, 2024; Zhang et al., 2025).

The most notable gains occurred in Grammar and Vocabulary Accuracy and Coherence and Cohesion, indicating students' responsiveness to layered feedback in areas requiring linguistic precision and overall text organisation (Zhang et al., 2025). Improvements in Vocabulary Range suggest that ChatGPT exposed learners to a wider repertoire of lexical choices, while teacher input refined appropriateness and contextual fit. In contrast, Task Fulfilment showed only marginal gains from ChatGPT feedback, reinforcing the importance of instructor mediation in clarifying communicative intent and task expectations (Najlaa et al., 2024; Zhang & Hyland, 2018).

These findings support Research Objective 1, and reaffirm that hybrid feedback is more effective than ChatGPT feedback alone. The combination of AI-generated and instructor-provided feedback enhanced both the accuracy and depth of revision, enabling students to produce writing that was not only more correct but also more coherent, purposeful, and contextually aligned.

The findings of this study provide empirical support for the Hybrid Feedback Model (see Figure 1), which conceptualises the writing process as a scaffolded progression within Vygotsky's Zone of Proximal Development (ZPD). The sequential improvement from ChatGPT-assisted to instructor-guided revisions mirrors the model's layered structure, where learners first engage in AI-supported revision within their ZPD before advancing toward deeper, instructor-mediated refinement. The marked gains observed in coherence, organisation, and linguistic accuracy demonstrate how each feedback stage activates different levels of cognitive engagement—moving from surface-level correction to conceptual reorganisation. This progression reinforces the model's assumption that effective writing development occurs through recursive feedback cycles, in which AI provides immediate linguistic scaffolding and instructor input consolidates higher-order learning. In this sense, the model not only guided the study's design but was also validated through the observed outcomes, underscoring its pedagogical relevance for integrating generative AI within process-oriented writing instruction.

Beyond these measurable improvements in writing performance, it was also essential to understand how students perceived the use of ChatGPT as a feedback tool. Therefore, the second research objective (RO2) examined students' attitudes, experiences, and preferences toward ChatGPT-mediated feedback to complement the quantitative findings.

RO2: ESL students' perceptions of ChatGPT as a writing feedback tool

The findings for Research Objective 2 are organized with quantitative results from closed-ended responses presented first, followed by qualitative insights derived from thematic analysis of open-ended responses.

STUDENTS' PRIOR KNOWLEDGE AND USAGE OF CHATGPT

The analysis of students' responses to Q4 (Figure 9) and Q5 (Figure 10) provides insight into their familiarity with ChatGPT prior to the intervention.

4. How often do you use ChatGPT for writing BEFORE today's class?

13 responses

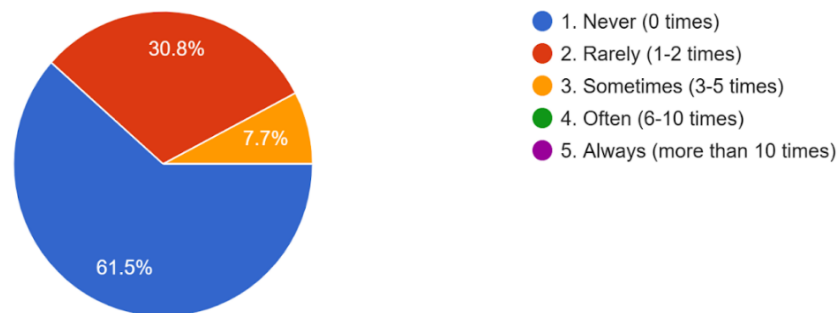


FIGURE 9. Frequency of ChatGPT usage before Intervention

Figure 9 shows that most students (61.5%) had never used ChatGPT for writing, with limited prior usage reported by 38.5%.

5. How well can you use ChatGPT for writing BEFORE today's class? Rate from 1 (not well) to 5 (very well).

13 responses

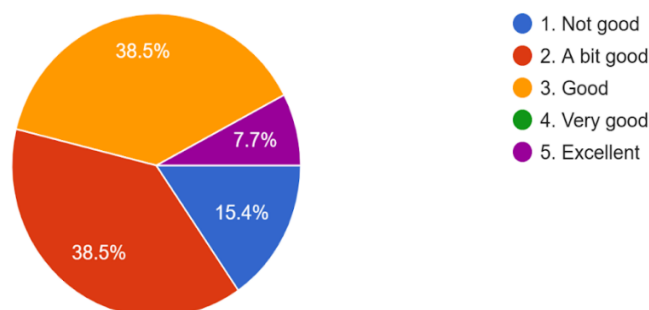


FIGURE 10. Students' Self-Rated Proficiency in using ChatGPT before Intervention

Meanwhile, Figure 10 shows that most students rated their ChatGPT proficiency as moderate (77%), with 38.5% selecting “Good” and 38.5% selecting “A bit good”, while fewer considered it very good (15.4%) or excellent (7.7%), highlighting the need for preparatory training.

PERCEIVED USEFULNESS FOR WRITING

Students’ perceptions of ChatGPT’s usefulness for writing are analysed through responses to Q7 (Figure 11) and Q12 (Figure 12).

7. ChatGPT helps students write essays and articles.

13 responses

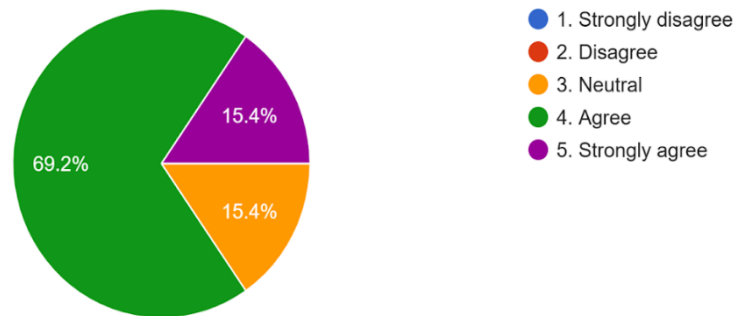


FIGURE 11. Perceived Usefulness of ChatGPT in Helping Students Write Essays and Articles

The majority of students (69.2%) agreed or strongly agreed that ChatGPT helps in writing essays and articles, indicating broad recognition of its utility for writing tasks, as shown in Figure 11.

12. ChatGPT gives students very good ideas for writing.

13 responses

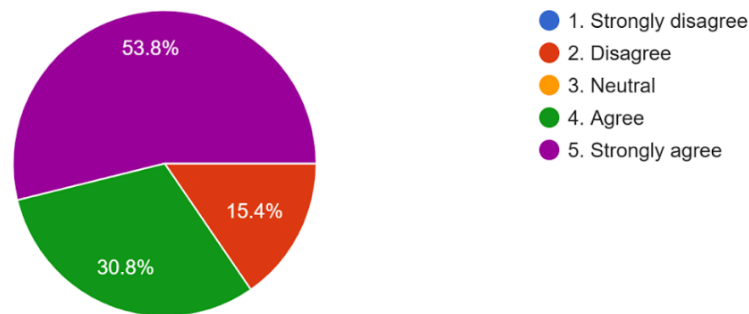


FIGURE 12. Perceived Effectiveness of ChatGPT in Generating Good Ideas for Writing

Additionally, Figure 12 demonstrates that over half of the students (53.8%) strongly agreed that ChatGPT provides very good ideas for writing, with an additional 30.8% agreeing.

BROADER PERCEPTIONS OF CHATGPT AS A LEARNING TOOL

Students' broader perceptions of ChatGPT were evaluated through Questions 9 (Figure 13) and 15 (Figure 14).

9. ChatGPT is a strong tool that helps people do more work.

13 responses

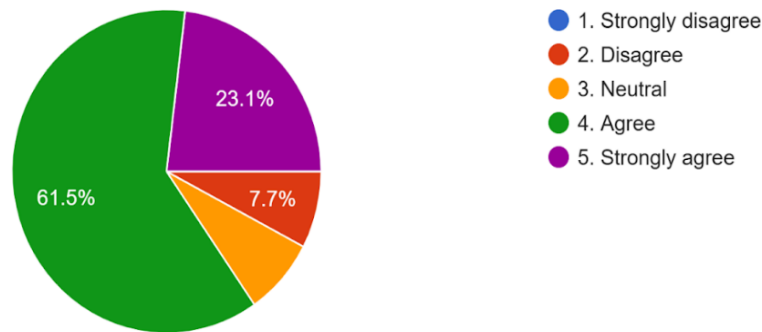


FIGURE 13. Perceptions of ChatGPT as a Tool to Enhance Productivity

Figure 13 shows that the majority of students agreed (61.5%) or strongly agreed (23.1%) that ChatGPT is a powerful tool for enhancing productivity, emphasizing its effectiveness in supporting academic tasks.

15. ChatGPT is important for universities because it helps students find answers.

13 responses

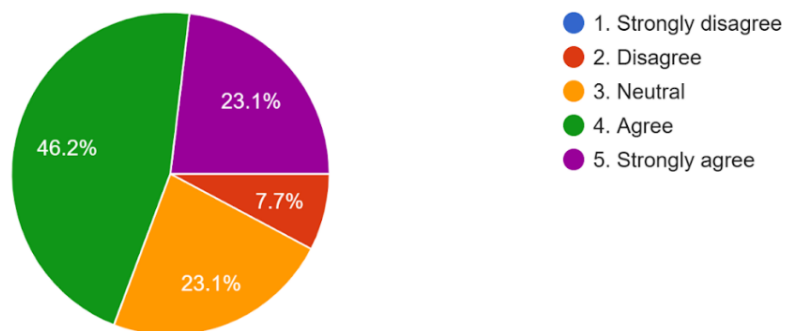


FIGURE 14. Perceptions of ChatGPT's Importance in University Learning Environments

Additionally, most students acknowledged ChatGPT's academic value, with 46.2% agreeing and 23.1% strongly agreeing that it aids universities as shown in Figure 14.

PERCEIVED BENEFITS OF CHATGPT FEEDBACK

Overall, students identified several benefits of ChatGPT feedback, particularly its ability to provide clear and accurate suggestions that helped them recognize mistakes and improve their writing skills. As one student reflected:

“It helps to improve my writing skills”.

Another student noted that it:

“improves my grammar skills.”

Many students also valued ChatGPT role in “expanding vocabulary”, describing it as generating ideas, and supporting self-revision through immediate, constructive feedback. Collectively, these comments indicate that students perceive ChatGPT as a supportive tool that enhance their autonomous learning, awareness and confidence in writing.

CHALLENGES IN USING CHATGPT FOR FEEDBACK

The most common challenge reported by students was internet connectivity issues, such as slow or unstable connections, which limited access to ChatGPT. One student remarked:

“They have a limit of uses and have to upgrade for more.”

Some students also mentioned the limited detail in the feedback provided by ChatGPT, noting that it occasionally lacked depth in addressing complex errors or broader writing issues. As one respondent explained:

“I have to tell ChatGPT to be more specific”

In addition, another student mentioned:

“Sometimes they don’t understand my question.”

A few respondents also cited occasional functionality problems, such as Internet issues and limited fthough some students experienced no major challenges at all.

PREFERENCE FOR FEEDBACK SOURCE: AI VS. HUMAN INSTRUCTOR

Most students expressed a preference for face-to-face feedback from lecturers, citing the benefits of real-time interaction and clearer explanations. One student commented:

“Even though ChatGPT is best to use for precise grammar checking and proofreading, one thing about it is sometimes it might give out details in a higher level English and sometimes people wouldn’t be able to understand well. But with a human instructor, it can guide us thoroughly based on our CEFR Level. Both have pros and cons.”

Another student commented:

“Human instructors are full of experience and they know the format well. Nevertheless, AI is also a great option since it’s much more efficient and faster compared to humans. However I’m still on HI (human instructor).”

However, some students appreciated the speed and accessibility of ChatGPT, particularly for quick revisions. As one student noted:

“It helps me give ideas and also corrects my mistakes which is very helpful.”

A minority also suggested combining both methods, leveraging ChatGPT’s immediacy and lecturers’ personalized guidance, as the ideal feedback approach. For instance, one student stressed that:

“I can compare the differences between a human instructor and AI tool and learn and gain my knowledge from both.”

SUMMARY OF FINDINGS AND ANALYSIS FOR RESEARCH OBJECTIVE 2

All in all, students tend to perceive ChatGPT as providing useful feedback for ESL writing, indicating broad recognition of its use in writing tasks. Many also highlighted ChatGPT's role in providing good ideas for writing, and it is a powerful tool to enhance productivity of academic tasks. This indicated ChatGPT's potential as a supportive tool for autonomous learning and self-reflection in university learning. This agrees with the finding that when students engage meaningfully with feedback, they become more aware of their writing processes and are able to revise independently (Zumbrunn et al., 2016).

Despite these benefits, there are several challenges identified in the use of ChatGPT in ESL writing classrooms. The most commonly reported issue was internet connectivity, which impeded access to the use of ChatGPT in ESL classrooms. In addition, some students perceived the feedback lacked depth and seemed superficial at times, especially when addressing more complex writing problems. Thus, this finding agrees that AI feedback alone can lack contextual understanding and task relevance (Najlaa et al., 2024).

Most students also expressed that they still prefer instructor input when comparing ChatGPT to human feedback, as instructor input provides an opportunity for clarification and prompt communication. Nevertheless, some appreciated the efficiency of being able to do quick checks and revisions on ChatGPT. Some students also opt for a hybrid approach, combining AI-generated feedback with personalised insights provided by human instructors. Again, this proves that instructor feedback remains essential due to its personalised nature that supports both linguistic accuracy and content quality (Taskiran & Goksel, 2022).

These findings suggest that, although ChatGPT can function as a beneficial feedback tool, it is not viewed as a complete replacement for instructor support. Instead, students appear to benefit the most when AI is integrated into a more hybrid feedback ecosystem that utilises both technology and human components. Therefore, this finding corresponds with research that emphasises the importance of feedback strategies in supporting learners' needs and preferences in ESL writing classrooms (Hyland & Hyland, 2001).

INTERPRETATION OF STUDENTS' PERCEPTIONS

Students' perceptions of ChatGPT as a feedback tool revealed a generally positive yet cautious stance toward AI-mediated learning. Most participants appreciated the immediacy, clarity, and accessibility of ChatGPT feedback, describing it as a useful starting point for revision. This aligns with Mahapatra (2024), who reported that ESL learners valued ChatGPT's instant response and supportive tone for building writing confidence. Similarly, Bok and Cho (2023) reported that students perceived ChatGPT as a useful revision partner that encouraged self-directed learning while maintaining reliance on teacher guidance for deeper content development. In this study, such perceptions suggest that ChatGPT fostered learners' sense of autonomy by enabling them to independently identify and correct surface-level issues, a process aligned with Vygotsky's (1978) Sociocultural Theory and the notion of scaffolded learning within the ZPD.

However, students also recognised the limitations of ChatGPT feedback, particularly its lack of contextual and task-specific understanding. These reservations echo the mixed findings reported in Shi and Aryadoust (2024), where students expressed uncertainty about the reliability and pedagogical depth of AI-generated feedback. Participants in this study frequently noted that ChatGPT's suggestions were generic or occasionally misaligned with the intended communicative purpose, reaffirming the need for instructor mediation to interpret, filter, and contextualise feedback. The preference expressed by many students for hybrid feedback—receiving initial input from ChatGPT followed by instructor clarification—illustrates an emerging model of complementary learning support.

Overall, these findings indicate that while learners recognise the benefits of AI for efficiency and exposure, they still value human feedback for its relevance, empathy, and explanatory power. The results therefore validate the Hybrid Written Feedback Model, suggesting that effective AI integration in ESL writing instruction depends not merely on technological availability but on pedagogical design that balances automation with human judgement.

NOVELTY OF THE STUDY

This study makes a novel contribution by examining students' engagement with ChatGPT as a feedback tool in an authentic ESL classroom, providing empirical support for the proposed Hybrid Written Feedback Model. The findings demonstrate that ChatGPT feedback enables learners to manage surface-level concerns such as grammar and vocabulary, while instructor feedback builds on this foundation to strengthen higher-order skills of coherence, organisation, and task fulfilment. This layered progression validates the model's integration of Sociocultural Theory (Vygotsky, 1978) and Process Writing Theory (Flower & Hayes, 1981), showing how scaffolded feedback within the ZPD supports recursive writing development.

The results extend existing theories of WCF by evidencing how sequential hybrid feedback facilitates both linguistic accuracy and higher-order revisions, reducing cognitive load at early stages and promoting deeper engagement at later ones. Methodologically, the integration of Likert-scale survey data with qualitative analyses illustrates how mixed-methods can capture not only writing outcomes but also learners' perceptions of AI-mediated feedback.

By situating these findings in a Southeast Asian ESL context, the study broadens global discussions on digital literacies and technology adoption, highlighting how cultural and institutional frameworks shape the pedagogical value of AI tools like ChatGPT. In doing so, it positions the Hybrid Written Feedback Model as a transferable theoretical lens for understanding human–AI collaboration in L2 writing, adaptable across diverse contexts.

CONCLUSION AND RECOMMENDATIONS

This study provides new evidence on how ESL learners engage with ChatGPT as a feedback tool within an authentic classroom context. The findings reveal that students valued ChatGPT for its immediacy and clarity, particularly in addressing surface-level issues such as grammar and vocabulary. At the same time, they relied heavily on instructor feedback for higher-order concerns, including coherence, organisation, and content development. These outcomes give empirical support to the Hybrid Written Feedback Model, which positions ChatGPT as scaffolding within the ZPD and places instructors as essential mediators for deeper revisions. In doing so, the study extends ongoing studies in WCF research and operationalises Process Writing Theory by showing how feedback can be layered to balance efficiency with depth.

The results also carry pedagogical implications. In the classroom, teachers may find it useful to integrate ChatGPT during early drafting stages, allowing students to manage surface-level revisions independently before receiving more targeted instructor guidance. Institutions have a role in ensuring that both teachers and students are equipped with digital literacy skills that enable them to interpret AI-generated feedback critically rather than accept it uncritically. As Qiong et al (2024) emphasised, sustainable digital transformation requires continuous training, institutional support, and reflective evaluation to enhance educators' digital competence. At the policy level, curriculum design should support hybrid feedback practices that view AI as a complement to teacher expertise, rather than as a substitute. At the policy level, curriculum design should support hybrid feedback practices that view AI as a complement to teacher expertise, rather than as a substitute. By embedding such blended approaches, educational systems can make better use of AI's strengths while preserving the irreplaceable value of human judgment in writing development. The study's findings resonate with recent educational perspectives that view digital transformation as central to pedagogical innovation. As Le Jinning et al (2025) argued, effective educational leadership now depends on integrating technological advancement with empathy-driven teaching to sustain meaningful learning in evolving digital environments.

While the design of the study ensured consistency, certain limitations should be acknowledged. The repeated use of a single writing task may have introduced task familiarity effects, and the fixed marking sequence may have unintentionally influenced rater expectations. The relatively small sample size ($n = 33$) also restricts the generalisability of findings. Future research could build on this work by varying writing prompts across phases, randomising the order of scoring, and recruiting larger, more diverse samples across institutions. It would also be valuable to explore whether learners of different CEFR proficiency levels engage differently with ChatGPT and instructor feedback, a question not fully addressed in this study but raised by the range of A1 to B2 learners in the sample. Finally, the study conceptualised Independent Mastery as the long-term goal of the Hybrid Model, but this stage was not empirically tested. Subsequent studies should examine how learners transfer hybrid feedback into sustained autonomy and long-term competence, perhaps through longitudinal or intervention-based designs.

In conclusion, this research advances theoretical and practical understanding of hybrid feedback by validating a model that combines AI and human input in a scaffolded sequence. It contributes to global discussions on AI-mediated learning by offering evidence from a Southeast Asian ESL context, while also providing clear pathways for teachers, institutions, and policymakers to integrate ChatGPT responsibly into writing pedagogy.

ACKNOWLEDGEMENT

The authors would like to express their gratitude to Universiti Teknologi MARA (UiTM) Cawangan Pulau Pinang for the institutional support provided throughout this study. Appreciation is also extended to the participating students for their cooperation and commitment during the data collection process. The constructive feedback from peer reviewers and colleagues has been invaluable in refining the quality of this paper.

REFERENCES

- Bok, E., & Cho, Y. (2023). Examining Korean EFL college students' experiences and perceptions of using ChatGPT as a writing revision tool. *Journal of English Teaching through Movies and Media*, 24(4). <https://doi.org/10.16875/stem.2023.24.4.15>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). Sage Publications.
- Ellis, R. (2009). Corrective feedback and teacher development. *L2 Journal*, 1(1), 3–18.
- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, 32(4), 365–387. <https://doi.org/10.2307/356600>
- Gopalan, M., Rosinger, K., & Ahn, J. B. (2020). Use of quasi-experimental research designs in education research: Growth, promise, and challenges. *Review of Research in Education*, 44(1), 218–243. <http://dx.doi.org/10.3102/0091732X20903302>
- Hyland, F., & Hyland, K. (2001). Sugaring the pill: Praise and criticism in written feedback. *Journal of Second Language Writing*, 10(3), 185–212. [https://doi.org/10.1016/S1060-3743\(01\)00038-8](https://doi.org/10.1016/S1060-3743(01)00038-8)
- Le Jinning, A., Aida Hanim, A. H., & Azlin Norhaini, M. (2025). Evolving Research Themes on Transformational Leadership in Basic Education (2000–2025). *Akademika*, 95(2), 275–292. <https://doi.org/10.17576/akad-2025-9502-14>
- Qiong, L.J., Muhammad Khair Noordin, Muhammad Afandi Azmi, Ahmad Nabil Md Nasir, & Mahyuddin Arsat. (2024). *Digital competencies for TVET educators: A framework for professional development*. *Akademika*, 94(3), 144–160. <https://doi.org/10.17576/akad-2024-9403-09>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11(9). <https://doi.org/10.1186/s40561-024-00295-9>
- Mathur, M., & Mahapatra, S. (2022). Impact of ePortfolio assessment as an instructional strategy on students' academic speaking skills: An experimental study. *CALL-EJ*, 23(3), 1–23. https://www.researchgate.net/publication/362394646_Impact_of_ePortfolio_Assessment_as_an_Instructional_Strategy_on_Students'_Academic_Speaking_Skills_An_Experimental_Study
- McCarthy, K. S., Roscoe, R. D., Allen, L. K., Likens, A. D., & McNamara, D. S. (2022). Automated writing evaluation: Does spelling and grammar feedback support high-quality writing and revision? *Assessing Writing*, 52, 100608. <https://doi.org/10.1016/j.asw.2022.100608>

- Miao, J., Chang, J., & Ma, L. (2023). Research trends of written corrective feedback in L2 writing: A bibliometric analysis. *Sage Open*, 13(1). <https://doi.org/10.1177/21582440221135172>
- Molin, F., Haelermans, C., Cabus, S., & Groot, W. (2020). The effect of feedback on metacognition - A randomized experiment using polling technology. *Computers & Education*, 149, 103885. <https://doi.org/10.1016/j.compedu.2020.103885>
- Najlaa, N. M. R., Aida, M., & Aileen, F. M. A. (2024). Evaluating automated grammar corrective feedback tools: A comparative study of Grammarly and QuillBot in ESL expository essays. *Akademika*, 94(2), 474–491. <https://doi.org/10.17576/akad-2024-9402-27>
- Nur Ashfaraliana, A. H., Mohamad, F., Md Johar, E., & Abdul Kadir, Z. (2024). Exploring the acceptance of ChatGPT as an assisting tool in academic writing among ESL undergraduate students. *International Journal of Research and Innovation in Social Science*, 8(10), 2886–2901. <https://doi.org/10.47772/IJRISS.2024.8100242>
- Raheem, B. R., Anjum, F., & Ghafar, Z. N. (2023). Exploring the profound impact of artificial intelligence applications (Quillbot, Grammarly, and ChatGPT) on English academic writing: A systematic review. *International Journal of Integrative Research*, 1(10), 599–622. <http://dx.doi.org/10.59890/ijir.v1i10.366>
- Shadiev, R., & Feng, Y. (2023). Using automated corrective feedback tools in language learning: A review study. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2022.2153145>
- Shi, H., & Aryadoust, V. (2024). A systematic review of AI-based automated written feedback research. *ReCALL*, 36(1), 1–23. <https://doi.org/10.1017/S0958344023000265>
- Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. *Assessing Writing*, 19(1), 51–65. <https://doi.org/10.1016/j.asw.2013.11.007>
- Storch, N., & Wigglesworth, G. (2010). Learners' processing, uptake, and retention of corrective feedback on writing: Case studies. *Studies in Second Language Acquisition*, 32(2), 303–334. <https://doi.org/10.1017/S0272263109990532>
- Taskiran, A., & Goksel, N. (2022). Automated feedback and teacher feedback: Writing achievement in learning English as a foreign language at a distance. *Turkish Online Journal of Distance Education*, 23(2), 108–123. <https://doi.org/10.17718/tojde.1096260>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Tran, T. T. T. (2025). Enhancing EFL Writing Revision Practices: The Impact of AI- and Teacher-Generated Feedback and Their Sequences. *Education Sciences*, 15(2), 232. <https://doi.org/10.3390/educsci15020232>
- Turner, E. (2023). Dialogic feedback and literary disciplinary knowledge in L2 writing instruction: How attitude to feedback influences academic achievement. *Research Papers in Education*, 38(1), 27–46. <https://doi.org/10.1080/02671522.2021.1941216>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. & Trans.). Harvard University Press. <https://doi.org/10.2307/j.ctvjf9vz4>
- Wang, L., Chen, X., Wang, C., Xu, L., Shadiev, R., & Li, Y. (2024). ChatGPT's capabilities in providing feedback on undergraduate students' argumentation: A case study. *Thinking Skills and Creativity*, 51, 101440. <https://doi.org/10.1016/j.tsc.2023.101440>
- Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511732997>

- Widad, M. A., Azman, C. M., Ghazali Yusri, A. R., & Rozaimi, J. (2024). Exploring the interconnections: Digital competence, teacher training and remote teaching. *Akademika*, 94(3), 108–123. <https://doi.org/10.17576/akad-2024-9403-07>
- Yao, Y., Yu, S., Zhu, X., Zhu, S., & Pang, W. (2024). Examining the mindsets toward students' writing competence and feedback-giving practices of Chinese university English writing teachers. *Studies in Educational Evaluation*, 80, 101318. <https://doi.org/10.1016/j.stueduc.2023.101318>
- Zhang, Z., & Hyland, K. (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36, 90–102. <https://doi.org/10.1016/j.asw.2018.02.004>
- Zhang, Z., Aubrey, S., Huang, X., & Chiu, T. K. F. (2025). The role of generative AI and hybrid feedback in improving L2 writing skills: A comparative study. *Innovation in Language Learning and Teaching*. <https://doi.org/10.1080/17501229.2025.2503890>
- Zimmerman, B. J., & Labuhn, A. S. (2012). Self-regulation of learning: Process approaches to personal development. In K. R. Harris, S. Graham, T. Urdan, C. B. McCormick, G. M. Sinatra, & J. Sweller (Eds.), *APA educational psychology handbook, Vol. 1: Theories, constructs, and critical issues* (pp. 399–425). American Psychological Association. <https://doi.org/10.1037/13273-014>
- Zumbrunn, S., Marrs, S., & Mewborn, C. (2016). Toward a better understanding of student perceptions of writing feedback: A mixed methods study. *Reading and Writing*, 29(2), 349–370. <https://doi.org/10.1007/s11145-015-9599-3>

BIODATA OF AUTHORS:

Aileen Farida Mohd Adam
Academy of Language Studies
UiTM Cawangan Pulau Pinang, Malaysia
Email: aileen@uitm.edu.my

Wan Najmiyyah Wan Md Adnan
Academy of Language Studies
UiTM Cawangan Terengganu, Malaysia
Email: wannajmiyyah@uitm.edu.my

Mohd Suhaimi Sulaiman (Corresponding author)
Faculty of Electrical Engineering
UiTM Cawangan Pulau Pinang, Malaysia
Email: shemi@uitm.edu.my