**Human versus Machine: How IELTS Learners Use AI Feedback on IELTS Writing Task 2 Essays and What They Miss**

*Ngoc Duyen Nguyen1*

**Abstract**

With the advent of artificial intelligence (AI) tools, such as ChatGPT, Grammarly, and QuillBot, the field of English language learning, especially in writing instruction, has been transformed. Although these platforms provide instant corrections and suggestions for improvement in language use and structure, it is ambiguous how effectively they address more complex writing requirements, particularly in task response and coherence. This small-scale classroom-based research examines how IELTS learners interpret and apply ChatGPT-generated feedback when revising their Writing Task 2 essays. Based on ChatGPT-revised essays, students’ reflections, and teacher-assigned IELTS scores, this study reveals that while AI tools assist in enhancing grammatical and lexical accuracy, students frequently misinterpret or overlook feedback regarding task response and coherence. This study also highlights the necessity for explicit AI-literacy training when preparing for IELTS, enabling students to critically evaluate the benefits and limitations of AI-generated feedback.

***Keywords*:** AI feedback, IELTS Writing, Chat GPT, Automated Writing Evaluation (AWE)

**1. Introduction**

Artificial intelligence (AI) tools have grown in popularity in English language learning, especially in writing. Platforms like Grammarly, Chat GPT and QuillBot offer real-time grammar, vocabulary, and structure recommendations. As a result, many English learners now depend on AI for self-editing and revision, particularly when preparing for important exams such as the International English Language Testing System (IELTS). Both independent learners and those enrolled in formal classes find these tools appealing due to the rapid, extensive, and frequently free assistance (Kurt & Kurt, 2024).

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1 Author, MA TESOL, MA Public Relations, American Degree Program, Duy Tan University, Danang, Vietnam; *Email: nguyenngocduyen3@dtu.edu.vn*

AI feedback is frequently used to supplement the limited amount of teacher time in IELTS preparation courses. This allows students to revise their essays more frequently. Yet, it remains unclear how effectively students understand and use the feedback these tools offer, especially when it comes to more abstract writing standards such as Task Response (TR) and Coherence and Cohesion (CC), which are crucial for success on IELTS Writing Task 2. In contrast to grammar or vocabulary corrections, current AI systems typically provide feedback that tends to be ambiguous or too broad with regard to logic, argument structure, or relevance (Koraishi, 2024).

In spite of increasing research on AI in education, few studies have investigated the actual learning experience, how students interact with AI feedback, which types of feedback they trust or dismiss, and whether their updated writing is really in alignment with IELTS band descriptors. In addition, there is a growing concern about overreliance on AI, misuse of paraphrasing tools, and the ethical issues of using machine-generated language in academic settings (Rizki & Anjarningsih, 2024).

This study examines how IELTS learners interpret and use AI-generated feedback from ChatGPT to improve their writing. Through a small classroom study, the research explores what types of feedback learners accept, misinterpret, or ignore, and how these choices influence their writing performance. The study suggests that IELTS classrooms should incorporate AI-literacy instruction for students to critically engage in and evaluate AI-generated recommendations.

**2. Literature Review**

***2.1. AI-based automated writing feedback***

According to Anderson (1982), feedback is significant in second language (L2) writing. Bitcherner et al. (2005) also asserted that among the different categories of feedback, error correction is the most commonly used. Many students believe that grammar correction from instructors enhances their writing skills. Nevertheless, it is time-consuming when providing feedback and therefore, more demanding and challenging for teachers to offer regular support. (Grimes & Warschauer, 2010). With the proliferation of technology, AI-based writing tools brings about a solution to alleviate teachers’ heavy workload and give students timely support. Due to the fact that today’s learners are digitally competent and appreciate immediate feedback (O’Neill & Russell, 2022), educators are under more pressure to offer effective and efficient responses.

The use of AI writing software, especially Automated Writing Evaluation (AWE) systems, has attracted significant interest in the study of L2 writing. These platforms apply artificial intelligence and natural language processing techniques to assess grammar and vocabulary. It has been consistently demonstrated across different studies that AWE elevates writing precision (Chang et al., 2021; Wilson & Roscoe, 2020; Zhai & Ma, 2022). Take the study of Cheng (2017) as an example, which found that Hong Kong undergraduates who received automated feedback outperformed those without. Similarly, Kurt and Kurt (2024) indicated how these tools are particularly beneficial in helping L2 learners to detect surface errors such as punctuation, choice of words, and sentence fragments. Additionally, Liao (2016) noted the gradual improvement in grammaticality across multiple rounds of revision; meanwhile, Rahman et al. (2022) and Gayed et al. (2022) highlighted more proficiency growth and fewer cognitive barriers to writing using AI-boosted tools. Moreover, it is proven in the research of Li et al. (2015) that US ESL learners made fewer grammatical, mechanical, and stylistic mistakes after using AWE, and they expressed highly positive attitudes toward its use.

Effective feedback is key to writing skill development. Bitchener and Ferris (2012) stressed the significance of being specific, actionable, and tailored to the needs of learners when providing feedback. However, this is not typically what AI-generated feedback is. The majority of AI programs available lack the ability to evaluate rhetorical strategies, clarity of arguments, or logical flow, which are identified as the higher-order concerns that distinguish strong academic writing (Koraishi, 2024). Consequently, learners may accurately revise their grammar or vocabulary while overlooking or misinterpreting suggestions in content relevance or structure areas. In a study which set out to determine students’ perspectives on machine feedback, Rizki and Anjarningsih (2024) demonstrated that while learners feel more confident in their writing after the implementation of AI feedback, they are also concerned with the tools’ oversimplicity and the shortage of contextual understanding. Several lines of evidence also raised concerns about the accuracy of feedback and how learners approach automated recommendations (Bai & Hu, 2017; Grimes & Warschauer, 2010). Ranalli et al. (2017) reported that despite the lack of precision of AWE feedback, students were still able to make corrections for errors based on that feedback in more than half of the cases. One study by Bai and Hu (2017) examined whether AWE software was precise when employed by 30 undergraduates in EFL writing classes. The results showed that they did not blindly accept automated feedback without question; instead, they considered meticulously which suggestions to adopt. This concluded that AWE should serve as a complement to, rather than a substitute for, human feedback, and further investigation should be conducted to explore the decision-making process on what automated feedback to use or ignore.

Furthermore, the heavy reliance on AI can raise ethical concerns. Students may become passive receivers of suggestions, and they may replace original ideas with AI-generated phrases without grasping the context (Kim et al., 2025). Not only does this limit learning, but it also calls into question authenticity, originality, and academic integrity.

***2.2. Writing feedback using ChatGPT***

The pervasiveness of AI writing tools has increasingly made it effortless for students to receive computer-provided feedback on their compositions. The tools, including ChatGPT, have transformed the practice of AWE by offering timely and personalised feedback (Gayed et al., 2022).

Thus far, a number of studies have revealed the instructional applications of ChatGPT in academic writing. A recent study by Mahapatra (2024) identified its role as a formative feedback tool for undergraduate ESL students and concluded that ChatGPT significantly improved students' writing skills. This suggests that ChatGPT can be employed as an effective source of feedback, which is particularly helpful in large classes where it is challenging for teachers to provide individual feedback. A broadly similar point has also been made by Dai et al. (2023), who studied the feasibility of ChatGPT for task-based feedback. The findings pointed out that ChatGPT could offer rich and thorough feedback that summarised student performance adequately, as well as surpassed the quality of feedback from human teachers in several aspects. This emphasises the potential of ChatGPT in enhancing students' learning approaches and writing performance. While AI-driven feedback systems have already shown promising effects on writing skills, adopting ChatGPT in writing instruction is still developing and necessitates empirical research (Barrot, 2023). Additionally, there is a growing need to examine how students’ perceptions of AI tools evolve as these technologies continue to improve.

***2.3. IELTS Writing and AI feedback***

IELTS Writing Task 2 is assessed across four criteria: Task Response (TR), Coherence and Cohesion (CC), Lexical Resource (LR), and Grammatical Range and Accuracy (GRA). While AI tools are generally helpful in addressing grammar and vocabulary (LR, GRA), they struggle with assessing TR and CC, the more abstract and interpretative features of writing.

Many recent studies highlight the growing interest in AI automated feedback for IELTS writing preparation. Grammarly and Criterion, for example, have been proven to elevate writing accuracy through a reduction in grammatical and lexical errors. (Wang et al., 2013; Li et al., 2015). IELTS test-takers who require concentrated writing practice can enjoy the benefits of instant, personalised feedback. Given the rise of AI-driven tools like ChatGPT, new possibilities for formative IELTS writing feedback have emerged. This view is supported by Mahapatra (2024) and Dai et al. (2023), who suggest that this tool is able to offer explicit, detailed feedback to aid learners in revising and refining their work effectively. In the same vein, Bui & Barrot (2024) stressed that ChatGPT is a supplementary tool, enabling students to draft their assignments and then improve clarity and accuracy.

On the other hand, Koraishi (2024) found that AI systems often over-rate TR and CC, especially when responses have complex sentence structures or use transitions. However, this can mask weak argumentation, irrelevant ideas, or a failure to fully address the question. TR, for instance, requires learners to respond to every aspect of the prompt with an articulated and evidenced stance, while AI tools lack the ability to reliably evaluate this, particularly when the responses are vague and formulaic. Similarly, CC is the element that pertains to logical flow and paragraphing, depending on the intuition of progression and unity by the reader rather than simply identifying cohesive devices. Therefore, learners who rely solely on AI suggestions may attain a delusive sense of competence in TR and CC, causing unsatisfactory outcomes in genuine test settings. Koraishi (2024) also found disparities between band scores produced by trained IELTS examiners and those generated by AI, particularly with regard to coherence and task achievement. These inconsistencies suggest that while AI can support the language learning process, learners should not be exclusively dependent on it as the only source of feedback for high-stakes test preparation like IELTS.

Despite its potential, ChatGPT’s specific contributions to IELTS preparation have not yet been fully explored, and further research on its impacts on all key IELTS scoring criteria (TR, CC, LR, and GRA) is necessary.

This study aims to fill these gaps by examining how IELTS learners interpret and apply ChatGPT-generated feedback, especially regarding TR and CC, how they compare it with teacher feedback, and how AI feedback influences actual performance and learner reflection in authentic IELTS writing tasks.

***Research questions:***

1. How do IELTS learners interpret and apply Chat GPT’s AI-generated writing feedback?
2. What types of feedback do they accept, ignore, or misinterpret?
3. Does AI feedback help improve performance in task response and coherence?
4. How do learners perceive the usefulness of AI versus teacher feedback?

**3. Methodology**

***3.1. Research design***

This study adopts a qualitative classroom-based research design supplemented with basic score comparison. It aims to explore how IELTS learners use AI-generated feedback in revising their essays, with a focus on Task Response (TR) and Coherence and Cohesion (CC). The goal is to understand not just the outcomes, but the learners’ reasoning and perceptions behind adopting or rejecting AI suggestions.

***3.2. Participants***

The study involves **12 upper-intermediate Vietnamese learners** (IELTS band 5.5–6.5 range), aged 17–21, enrolled in an IELTS course at Troy University Danang *(See Appendix A for full demographic information)*. All participants have some prior exposure to AI tools and will be given the choice to use either Grammarly, ChatGPT, or QuillBot for revision tasks. Participants were provided a consent form *(See Appendix B)*.

***3.3. Data collection***

Data collection occurs over three phases.

In phase 1, 12 students were required to write an essay to answer an IELTS Task 2 question individually under test conditions, without any AI assistance. Each essay was labelled from 1 to 12.

*Essay question:*

***Some people say children should not have mobile phones until they are teenagers, while others argue they should learn how to use them from a young age.******Discuss both views and give your opinion.***

*Give reasons for your answer and include any relevant examples from your own knowledge or experience.*

*Write at least 250 words.*

In phase 2, a brief instructional session (20 minutes) was conducted, where the researcher explained how to interpret ChatGPT feedback, modelled how to revise based on it, and encouraged students to verify ChatGPT suggestions using IELTS band descriptors. The students then revised the same essay at home using ChatGPT with a fixed prompt designed to elicit consistent and pedagogically relevant feedback. While ChatGPT responses may vary slightly across sessions, using a fixed prompt and controlled writing samples reduced variability and enhanced comparability across feedback instances. The prompt was as follows:

*“Please give detailed feedback on this IELTS Writing Task 2 essay using the official IELTS Writing Band Descriptors.”*

This prompt was piloted and refined during a pre-study phase to ensure that ChatGPT-generated feedback aligned with assessment criteria familiar to learners and researchers.

Students were then instructed to save and submit both the original and the revised versions. Each revised essay was labelled from 1 to 12 according to the order of pre-edited essays.

In phase 3, they completed a feedback form *(see Appendix C)* indicating what changes they made and why to explore their experiences, challenges, and perceptions of AI feedback.

Additionally, the pre- and post-revision essays were scored by an IELTS-certified teacher using the official IELTS band descriptors, with a focus on TR and CC.

***3.4. Data analysis***

The data is analysed in three stages. Firstly, teacher scores for TR and CC were compared between the original and ChatGPT-revised essays to identify gains or regressions. Then, student reflections were coded thematically to identify patterns in how learners interpret and apply ChatGPT feedback (e.g., accepted vs. rejected suggestions, surface vs. content revisions). After that, a subset of ChatGPT feedback was categorised by type, grammar, vocabulary, organisation, and logic, to determine what learners focus on and what they overlook.

To enhance the trustworthiness and pedagogical relevance of the ChatGPT-generated feedback, a validation procedure was conducted using expert human judgment. A stratified sample of 20 feedback instances (5 credits for each IELTS Writing Band Descriptors: TR, CC, LR and GRA) was independently reviewed by two certified IELTS raters with extensive TESOL experience. The purpose was to assess the accuracy and appropriateness of ChatGPT’s comments on the four IELTS Writing Band Descriptors. Each rater coded the ChatGPT-generated feedback based on whether it accurately addressed the learner’s strengths and weaknesses in relation to the relevant descriptor. Feedback was categorised using a three-point scale:

* Accurate (A): The feedback clearly and correctly identified issues or strengths aligned with IELTS descriptors.
* Partially Accurate (P): The feedback was somewhat relevant but lacked depth or misrepresented certain aspects.
* Inaccurate (I): The feedback was irrelevant, vague, or contradicted accepted IELTS assessment principles.

Inter-rater agreement was calculated using both percentage agreement and Cohen’s Kappa to account for chance agreement. The results are presented in Table 1.

**Table 1.**

*Inter-Rater Agreement on ChatGPT Feedback Accuracy (n = 23 credits × 2 raters = 46 total judgments)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **IELTS Band Descriptors** | **Rater 1**  **(A / P / I)** | **Rater 1**  **(A / P / I)** | **Agreement (%)** | **Cohen’s Kappa** |
| Task Response (TR) | 18 / 4 / 1 | 17 / 5 / 1 | 86.9% | 0.76 |
| Coherence and Cohesion (CC) | 19 / 3 / 1 | 18 / 4 / 1 | 87% | 0.75 |
| Lexical Resource (LR) | 17 / 5 / 1 | 16 / 6 / 1 | 82.6% | 0.71 |
| Grammatical Range & Accuracy (GRA) | 18 / 4 / 1 | 17 / 5 / 1 | 86.9% | 0.74 |
| **Overall** |  |  | **85.9%** | **0.74** |

The inter-rater agreement (κ = 0.74) indicates substantial agreement between the two raters (Landis & Koch, 1977) *(see Appendix D)* across all four descriptors, showing that ChatGPT feedback was generally accurate and pedagogically sound. Discrepancies were discussed and resolved to reach full consensus. Insights from the validation phase informed several revisions to the prompt used with ChatGPT. For instance, generic suggestions such as *"improve grammar"* were clarified to elicit feedback targeting specific patterns *("increase the use of complex sentence structures"* or *"review subject-verb agreement in conditional sentences")*. These refinements improved the pedagogical clarity and consistency of ChatGPT-generated feedback for subsequent learners.

ChatGPT feedback was coded thematically using an inductive coding strategy, in which codes are developed directly from the data collected from participants. The feedback was then compared with student uptake and post-feedback writing performance, to assess whether changes aligned with ChatGPT's suggestions.

**4. Results**

This section presents the outcomes of a comparative analysis of IELTS Writing Task 2 essays before and after learners engaged with ChatGPT’s AI-generated feedback. The results are organised around the four guiding research questions. A sample of 12 learners (n = 12) submitted paired writing tasks, which were analysed both qualitatively and quantitatively.

***4.1 Interpretation and application of AI feedback***

Analysis of student-revised essays and reflection sheets showed that learners mainly applied low-level corrections, particularly grammar and vocabulary. ChatGPT’s feedback often included broader suggestions, such as rephrasing ideas or reordering paragraphs, but students rarely implemented these. In 12 essay pairs, students applied over 230 surface-level edits, compared to just 31 higher-order revisions.

***Table 2:***

*Types and number of feedback by ChatGPT*

|  |  |  |
| --- | --- | --- |
| **Feedback Type** | **Strategy** | **Occurrences** |
| **Grammar** | Tense correction | 52 |
| Subject–verb agreement | 37 |
| Article use | 21 |
| Word order | 19 |
| **Lexical Resource** | Word substitution | 44 |
| Synonym replacement | 22 |
| Formal word adjustment | 13 |
| **Coherence & Cohesion** | Add linking device | 11 |
| Reorder sentence | 6 |
| Add topic sentence | 3 |
| **Task Response** | Add example or explanation | 5 |
| Clarify main idea | 3 |

All 12 learners demonstrated clear attempts to apply ChatGPT’s feedback. The feedback was generally interpreted correctly, especially when it involved explicit grammatical corrections or vocabulary substitution. The majority of learners incorporated revised phrases, corrected structures, and improved paragraph transitions.

However, the depth of application of ChatGPT’s feedback varied considerably across learners. While surface-level suggestions, such as correcting grammatical errors, adjusting word forms, and replacing informal vocabulary, were almost universally accepted, deeper revisions related to content structure and critical thinking were applied inconsistently. For instance, in the revised version 1, the learner successfully improved sentence clarity, eliminated run-on sentences, and adjusted tone to better suit an academic context. Phrases such as *“phones are not good”* were upgraded to more formal constructions like *“mobile phones can negatively affect students,”* and casual expressions were replaced with appropriate academic vocabulary. However, the core content of the argument remained underdeveloped, and the learner did not address the deeper issue of balancing both perspectives with more nuanced examples of counterarguments.

**Example 1: Student 1 – Essay pair 1**

**Original sentence:** *“Phones are not good for kids because they just play games.”*

**ChatGPT feedback (LR):** *“Avoid vague or informal expressions like ‘phones are not good.’ Try: ‘Mobile phones can negatively affect students’ concentration or social skills.’”*

**Revised sentence:** *“Mobile phones can negatively affect children’s concentration and limit their outdoor activities.”*

**ChatGPT feedback (TR):** *“Can you elaborate on how phones impact social development or provide an example of excessive screen use?”*

**Student revision:**  
*None added. The next sentence was: “That’s why they should not have phones too early.”*

A similar pattern emerged in the revised version 10, featuring more accurate grammar, smoother transitions, and enhanced lexical range, yet the learner retained vague reasoning and did not elaborate on key claims. For example, the idea that students lack study skills at university was briefly mentioned but not expanded with clear supporting evidence or real-life illustrations. Likewise, in the revised version 5, it showed much clearer paragraph organisation and improved vocabulary (*“advantages”* instead of *“strongest point”*), yet the arguments remained simplistic and occasionally generic. This indicates that while learners are generally adept at implementing mechanical and linguistic changes, they often overlook or are unsure how to apply more abstract feedback, such as refining thesis focus, developing arguments logically, or offering concrete examples to support claims. Thus, ChatGPT’s feedback appears to be most effective when addressing clarity and correctness, but less so in fostering critical thinking and content expansion. This suggests that while learners are comfortable applying mechanical and linguistic improvements, they may struggle with or overlook more abstract or content-related feedback.

**Example 2: Student 5 – Essay pair 5**

**Original sentence:** *“The strongest point is kids can contact their parents if something happens.”*

**ChatGPT feedback (LR):** *“The phrase ‘strongest point’ is too informal. Consider replacing it with ‘main advantage’ or ‘primary benefit’ to suit academic tone.”*

**Revised sentence:** *“The main advantage is that children can use mobile phones to contact their parents if something happens.”*

**ChatGPT Feedback (TR):** *“Try to balance your discussion. Can you include a possible downside to early phone use or show that not all children use them responsibly?”*

**Student revision:** The student added a short clause: *“But some kids can use phones for bad things.”*

**4.2 Types of feedback: accepted, ignored, or misinterpreted**

The analysis of feedback usage is summarised in Table 3 below.

**Table 3.**

*Learner Response to ChatGPT Feedback (n = 12)*

|  |  |  |  |
| --- | --- | --- | --- |
| **Feedback type** | **Accepted (%)** | **Ignored (%)** | **Misinterpreted (%)** |
| Grammar corrections | 100% | 0% | 0% |
| Vocabulary enhancement | 86% | 14% | 0% |
| Sentence structure | 100% | 0% | 0% |
| Idea development | 43% | 57% | 0% |
| Logical progression | 71% | 29% | 0% |
| Use of transitions | 86% | 0% | 14% |
| Tone and academic style | 71% | 14% | 14% |

Learners were most responsive to grammatical, vocabulary, and sentence structure suggestions, indicating that such changes are easier to interpret and implement. However, more complex feedback involving idea development and argument logic was often ignored or only partially understood. In some cases, learners inserted transitions like *“On the one hand”* or *“In conclusion”* without improving actual argumentative flow, reflecting partial misinterpretation.

***4.3 Improvement in Task Response and Coherence***

To quantify writing improvements, each pair of essays was rated based on IELTS band descriptors across four criteria: Task Response (TR), Coherence & Cohesion (CC), Lexical Resource (LR), and Grammatical Range and Accuracy (GRA).

**Table 4.**

*Average IELTS band scores before and after ChatGPT feedback*

|  |  |  |  |
| --- | --- | --- | --- |
| **Criterion** | **Before** | **After** | **Band increase** |
| Task Response (TR) | 5.0 | 5.7 | +0.7 |
| Coherence & Cohesion (CC) | 5.5 | 6.3 | +0.8 |
| Lexical Resource (LR) | 5.3 | 6.2 | +0.9 |
| Grammar Range & Accuracy (GRA) | 5.0 | 6.5 | +1.5 |

As illustrated in the table 4 above, the most significant gain occurred in GRA, with an average improvement of +1.5 bands. This aligns with learners’ consistent acceptance of grammar-related feedback.

LR and CC also improved notably, indicating that learners were able to integrate more precise vocabulary and organise ideas more clearly.

However, TR exhibited only moderate improvement, suggesting that while ChatGPT feedback improved presentation and language, it was less effective in helping learners strengthen content and argument development.

***4.4. Learner perceptions: AI vs. teacher feedback***

Learners’ reflections on the usefulness of AI-generated feedback, particularly from ChatGPT, revealed a generally positive attitude toward its application in academic writing. Across the three student feedback forms analysed, all learners acknowledged that ChatGPT contributed significantly to improving their essays. However, their views also revealed a nuanced understanding of the limitations and ideal use of AI feedback in contrast to teacher feedback.

*4.4.1. Perceived usefulness of ChatGPT*

Three students reported that ChatGPT was helpful in identifying grammatical errors, suggesting better vocabulary, and enhancing fluency and clarity. For example, student 2 shared that ChatGPT provided specific suggestions to improve sentence structure and word choice, leading to a more academic and cohesive writing style. Similarly, student 4 highlighted that the tool helped detect mistakes in using tenses and vocabulary, and it offered smoother alternatives to enhance readability. Student 3 noted that ChatGPT helped link ideas more clearly and expand her arguments appropriately.

Beyond surface-level edits, some learners appreciated how ChatGPT supported idea development. Student 5 reported that it provided clear guidance on how to avoid redundancy and maintain logical progression between points. Another commented that ChatGPT offered assistance in framing a better introduction and conclusion, which are areas many learners typically find difficult to construct.

*4.4.2. Limitations and cautions when adapting ChatGPT feedback*

Despite these strengths, learners expressed important reservations. A recurring concern was ChatGPT inconsistency and reliability. For instance, student 5 and student 6 mentioned that when they used three different AI tools to estimate their band score, each returned a different result, leading them to conclude that AI feedback should be used for reference only, not as an authoritative assessment tool. They also noted that ChatGPT could miss certain types of mistakes, especially more subtle ones.

Other learners highlighted the lack of context-awareness in ChatGPT responses. For example, student 10 explained that while the suggestions were mostly reliable, they had to double-check ChatGPT’s advice against the prompt to ensure it remained relevant to the task. Meanwhile, student 9 emphasised the importance of not becoming dependent on ChatGPT, warning that overuse may hinder one’s ability to think critically and write independently.

*4.4.3. Comparison with teacher feedback*

Although direct teacher feedback was not provided in the same format for comparison, the students’ comments implied a distinction. ChatGPT was often seen as a convenient tool for grammar and vocabulary, offering instantaneous, multiple suggestions that are especially helpful for independent learners. However, learners also acknowledged that human teachers are still essential, especially when it comes to evaluating task achievement, cultural appropriateness, and deep content quality.

For example, while ChatGPT can suggest alternative phrasings or point out awkward expressions, it might not always recognise subtle errors in argument logic or relevance to the IELTS prompt. This suggests a clear preference for balancing AI suggestions with expert human judgment.

*Student 12 commented:* *“ChatGPT helps with fluency, but my teacher helped me organise ideas better and avoid going off-topic.”*

*Student 11 noted: “Although ChatGPT is useful, feedback from ChatGPT is only for reference.”*

*Student 4 commented:* *"My teacher gave feedback that was more focused on IELTS scoring. ChatGPT was more like general writing help."*

*4.4.4. Recommendations from learners*

Learners proposed several enhancements to improve the educational value of AI-generated feedback, particularly when using ChatGPT for IELTS writing practice. One common suggestion was the inclusion of simplified explanations to clarify why certain revisions are necessary. Instead of only providing the corrected version, learners wished to understand the reasoning behind the changes so they could internalise the rules and apply them independently in future writing. Additionally, students expressed the need for examples tailored to their English proficiency level, arguing that sample sentences or model paragraphs adjusted to their current ability would make the feedback more relatable and easier to follow.

Another highly requested feature was the ability to view before-and-after comparisons, which would allow learners to clearly see what had been changed and understand how the revised version improved on the original. This would also help reinforce grammar rules and stylistic improvements in a more visual and memorable way. Finally, learners emphasised the importance of making AI feedback more sensitive to the specific task requirements of IELTS prompts, such as clearly addressing all parts of the question, maintaining relevance, and avoiding off-topic ideas, areas where human teachers tend to provide more accurate judgment.

These thoughtful suggestions demonstrate that while learners appreciate the convenience and usefulness of ChatGPT, they are also aware of its current limitations. Their feedback reflects a desire not just to receive corrections, but to learn from the process, indicating that many students see AI as a powerful supplement, not a replacement, for teacher guidance.

**5. Discussion**

This study examined how IELTS learners interpret and apply ChatGPT-generated feedback, the types of feedback they respond to, the extent to which ChatGPT improves task response and coherence, and how learners perceive AI feedback in comparison to teacher feedback. The findings highlight the nuanced role that tools like ChatGPT can play in writing instruction, particularly for learners preparing for high-stakes assessments such as the IELTS exam.

***5.1. Learners prefer surface-level feedback: alignment with previous AWE studies***

Consistent with previous literature (e.g., Kurt & Kurt, 2024; Liao, 2016), the learners in this study primarily applied surface-level corrections, especially grammar and vocabulary-related changes. Over 230 grammar and lexical edits were implemented across 12 students, compared to only 31 higher-order revisions related to content, logic, or structure. This pattern supports findings from Li et al. (2015) and Wilson & Roscoe (2020), who noted that AWE tools are particularly effective at helping students correct sentence-level errors but less so at guiding deeper content development.

While ChatGPT provided feedback on broader rhetorical elements such as idea expansion and coherence, students rarely applied this feedback. This aligns with Koraishi (2024) and Bitchener and Ferris (2012), who argue that effective feedback must be actionable and tailored. Most current AI systems fall short in this regard, offering suggestions that lack instructional depth or contextual nuance, an issue raised by Rizki & Anjarningsih (2024) and confirmed in our findings. Students in this study often misunderstood or ignored higher-order feedback, particularly when it required them to evaluate argument logic or fully address all parts of the IELTS prompt.

For example, learners successfully refined tone and fluency but failed to address underdeveloped arguments. Similarly, some of the revised essays featured enhanced vocabulary and transitions, yet retained vague and unsupported reasoning. These cases reflect a broader trend: while learners are proficient in implementing surface-level improvements, they may lack the critical thinking or metacognitive strategies needed to evaluate and respond to more complex guidance.

***5.2. Interpretation and selective uptake: the learner’s role in feedback use***

The learners’ selective engagement with ChatGPT feedback mirrors previous findings by Bai & Hu (2017), who noted that L2 writers do not accept automated feedback uncritically. In this study, while all learners accepted grammar corrections (100%), only 43% applied feedback related to idea development. In several cases, learners inserted transition phrases such as *“On the one hand”* or *“In conclusion”* without enhancing logical progression, indicating a superficial application of rhetorical markers. This echoes Koraishi’s (2024) concern that AI-generated coherence may be formulaic and fail to reflect true cohesion or content unity.

Learners demonstrated the metacognitive ability to distinguish between helpful and questionable suggestions, as also reported in Ranalli et al. (2017). For instance, some students reported discarding ChatGPT’s suggestions when they contradicted their understanding of the task, especially in Task Response (TR) areas, a scoring component where AI often fails to detect off-topic or underdeveloped ideas (Koraishi, 2024).

***5.3. Effectiveness in improving band performance***

The IELTS scoring analysis revealed clear improvements in Grammatical Range and Accuracy (+1.5 bands) and Lexical Resource (+0.9 bands), confirming AI’s strength in refining form and style (Chang et al., 2021; Mahapatra, 2024). These improvements are in line with Gayed et al. (2022) and Zhai & Ma (2022), who documented improved sentence accuracy and vocabulary precision through AI-assisted revision.

However, improvements in Task Response (+0.7) and Coherence and Cohesion (+0.8) were modest, likely because these categories require critical thinking, task interpretation, and logical structuring, which generative AI like ChatGPT still struggles to assess effectively (Koraishi, 2024; Bui & Barrot, 2024). As noted in the literature, TR and CC require not just cohesive devices, but meaningful progression and engagement with all aspects of the question, skills that AI cannot reliably teach or evaluate (Li et al., 2015; Shang, 2019).

***5.4. Learner perceptions: opportunities and cautions***

Echoing findings from Barrot (2023) and Mahapatra (2024), most learners in this study reported positive experiences with ChatGPT, citing its helpfulness in fixing grammar and vocabulary and improving clarity. Students appreciated the tool’s immediacy and accessibility, consistent with O’Neill & Russell’s (2022) point that digital-native learners value instant feedback.

However, they also expressed caution about overreliance and the limitations of AI-generated feedback. For instance, students highlighted inconsistencies in band score estimations between different AI tools, and several learners were sceptical of ChatGPT’s ability to assess task relevance, concerns also documented by Bai & Hu (2017) and Kim et al. (2025). Some learners warned that excessive use of AI may inhibit independent thinking, reinforcing ethical concerns around originality and cognitive engagement.

These findings support the growing consensus that AI tools like ChatGPT are best used as supplements, not replacements, for human feedback (Shang, 2019; Bitchener & Ferris, 2012). While AI can efficiently correct surface errors and enhance fluency, teacher feedback remains critical for guiding content development, task interpretation, and argument coherence.

***5.5. Implications for practice and AI design***

The data confirms that ChatGPT offers valuable support for micro-level writing development, such as grammar correction, lexical refinement, and sentence-level clarity. However, its ability to guide macro-level writing strategies, such as structuring cohesive arguments or fully addressing the IELTS prompt, is limited. This limitation has important implications for how AI tools are used in IELTS preparation.

For classroom practice, IELTS instructors should be encouraged to implement AI-literacy tasks that help students critically evaluate, adapt, and discuss AI-generated feedback. For example, students could compare AI feedback with peer and teacher feedback, reflect on which suggestions they applied and why, or revise their writing in stages with guided prompts; for example,“Which of ChatGPT’s ideas do you agree with? Which ones need clarification?”. Teachers could also integrate AI-reflection journals or “think-aloud” revision logs as part of portfolio assessment to promote deeper engagement with both surface- and content-level feedback.

For teacher training and curriculum design, these findings support a shift toward a blended feedback model that positions AI as a supplement, not a replacement, for teacher feedback. Instructors may need targeted professional development on how to scaffold AI use, monitor learner misconceptions, and provide higher-order support that aligns with IELTS descriptors such as TR and CC. For example, teachers could design writing lessons where AI is used only for first drafts or for language polishing, while content and structure are addressed through peer review or class discussion.

For AI tool development, learners in this study suggested useful enhancements, including simpler explanations, before/after writing examples, and clearer alignment with learner proficiency levels. These ideas reinforce recommendations by Barrot (2023) and Dai et al. (2023), who argue for pedagogically informed AI design. Future tools should include level-adaptive feedback, visual or scaffolded suggestions for structural improvements, and interactive features such as clickable explanations, genre-specific templates, or customizable tone and formality settings. Importantly, AI systems should be transparent about their limitations, encouraging users to consult teachers or human experts when needed.

At the policy level, language education programs and assessment bodies should develop guidelines for ethical and pedagogical AI integration. Institutions could include AI-use protocols in writing curriculum, provide access to approved AI tools, and create evaluation rubrics that reward students’ ability to interpret and apply feedback critically rather than simply accepting suggestions.

**6. Conclusion**

This study explored how IELTS learners perceive, utilise, and regard AI-generated feedback, particularly from ChatGPT, when revising their Writing Task 2 essays. The findings offer insight into student engagement with automated writing support and reveal both the potential and the limitations of AI for academic writing improvement.

The results showed that while learners consistently applied low-level feedback, especially related to grammar, vocabulary, and sentence structure, they were far less likely to act on higher-order suggestions such as improving task response, elaborating arguments, or refining essay structure. This pattern was evident in the analysis of 12 essay pairs, where surface-level edits vastly outnumbered deeper content changes. Quantitative scoring confirmed that the most significant gains occurred in grammatical range and lexical resource, while improvements in task response were more modest.

Learner reflections showed that ChatGPT is widely appreciated for its ability to provide quick, accessible, and constructive feedback, particularly in contexts where teacher input is limited. However, students also expressed valid concerns about its reliability, task sensitivity, and limitations in assessing idea development or argument relevance. Many learners viewed ChatGPT as a useful complement to, but not a substitute for, teacher feedback, which they regarded as more comprehensive and aligned with exam criteria.

Importantly, learners expressed a desire for AI tools to evolve beyond correction and offer more instructional value through clearer explanations, level-appropriate examples, and features that help users learn from their mistakes. These insights suggest that with thoughtful design and integration into classroom practices, AI can serve not just as a grammar checker but as a meaningful learning companion that helps students build independent writing skills.

Future studies could expand on this work by comparing AI and teacher feedback side by side on the same essays or by examining long-term writing development across multiple assignments. It would also be valuable to explore how learners of different proficiency levels or learning contexts (e.g., online versus in-class) respond to AI-generated feedback over time. Lastly, as AI continues to evolve, ongoing evaluation of its pedagogical effectiveness and ethical use in educational settings remains essential.

**References**

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369-406. [https://doi.org/10.1037/0033-295X.89.4.369](https://psycnet.apa.org/doi/10.1037/0033-295X.89.4.369)

Bai, L., & Hu, G. (2017). In the face of fallible AWE feedback: How do students respond?. *Educational Psychology*, *37*(1), 67-81. <https://doi.org/10.1080/01443410.2016.1223275>

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. Assessment in Writing, 57, 100745. <https://doi.org/10.1016/j.asw.2023.100745>

Bui, N. M., & Barrot, J. S. (2024). ChatGPT as an automated essay scoring tool in the writing classrooms: how it compares with human scoring. *Education and Information Technologies*, 1-18. <https://doi.org/10.1007/s10639-024-12891-w>

Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing*. Routledge.

Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of second language writing*, *14*(3), 191-205. <https://doi.org/10.1016/j.jslw.2005.08.001>

Chang, T. S., Li, Y., Huang, H. W., & Whitfield, B. (2021, March). Exploring EFL students' writing performance and their acceptance of AI-based automated writing feedback. In *Proceedings of the 2021 2nd International Conference on Education Development and Studies* (pp. 31-35). <https://doi.org/10.1145/3459043.3459065>

Chen, X., Zhou, Z., & Prado, M. (2025). ChatGPT-3.5 as an automatic scoring system and feedback provider in IELTS exams. *International Journal of Assessment Tools in Education*, *12*(1), 62-77. <https://doi.org/10.21449/ijate.1496193>

Cheng, G. (2017). The impact of online automated feedback on students' reflective journal writing in an EFL course. *The Internet and Higher Education*, *34*, 18-27. <https://doi.org/10.1016/j.iheduc.2017.04.002>

Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y. S., Gašević, D., & Chen, G. (2023, July). Can large language models provide feedback to students? A case study on ChatGPT. In *2023 IEEE International Conference on advanced learning technologies (ICALT)* (pp. 323-325). IEEE.

Koraishi, O. (2024). The intersection of AI and language assessment: A study on the reliability of ChatGPT in grading IELTS Writing Task 2. *Language Teaching Research Quarterly*, 43, 22–42. <https://doi.org/10.32038/ltrq.2024.43.02>

Gayed, M., Carlon, J., Oriola, M., & Cross, S. (2022). Exploring an AI-based writing assistant’s impact on English language learners. *Computer & Education: Artificial Intelligence, 3*, 100055. https://doi.org/10.1016/j.caeai.2022.100055

Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. *The Journal of Technology, Learning and Assessment*, *8*(6).

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

Kurt, G., & Kurt, Y. (2024). Enhancing L2 writing skills: ChatGPT as an automated feedback tool. *Journal of Information Technology Education: Research*, *23*, 024. DOI: 10.28945/5370

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.

Leki, I. (1991). The preferences of ESL students for error correction in college‐level writing classes. *Foreign language annals*, *24*(3), 203-218.

Li. J., Link, S., & Hegelheimer, V. (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing, 27*, 1-18. https://doi.org/10.1016/j.jslw.2014.10.004

Liao, H.-C. (2016). Enhancing the grammatical accuracy of EFL writing by using an AWE-assisted process approach. *System, 62,* 77-92. <https://doi.org/10.1016/j.system.2016.02.007>

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

O'Neill, R., & Russell, A. (2019). Stop! Grammar time: University students’ perceptions of the automated feedback program Grammarly. *Australasian Journal of Educational Technology*, *35*(1). <https://doi.org/10.14742/ajet.3795>

Rahman, N., Zulkornain, H., & Hamzah, H. (2022). Exploring artificial intelligence using automated writing evaluation for writing skills. *Environment-Behaviour Proceedings Journal, 7*, 547-553. <https://doi.org/10.21834/ebpj.v7iSI9.4304>

Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: Investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology*, *37*(1), 8-25. <https://doi.org/10.1080/01443410.2015.1136407>

Rizki, M., & Anjarningsih, H. Y. (2024). The use of AI agents to help teach IELTS Writing Task 2: A narrative inquiry. *Jurnal Cahaya Mandalika*, 5(2), 985–999.<https://doi.org/10.36312/jcm.v5i2.3865>

Shang, H. F. (2022). Exploring online peer feedback and automated corrective feedback on EFL writing performance. *Interactive Learning Environments*, *30*(1), 4-16. <https://doi.org/10.1080/10494820.2019.1629601>

Wang, Y.-J., Shang, H.-K., & Briody, P. (2013). Exploring the impact of using automated writing evaluation on English as a foreign language university students’ writing. *Computer Assisted Language Learning, 26*(3), 234-257. <https://doi.org/10.1080/09588221.2012.655300>

Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research*, *58*(1), 87-125. [https://doi.org/10.1177/07356331198307](https://doi.org/10.1177/0735633119830764)

Zhai, N., & Ma, X. (2022). Automated writing evaluation (AWE) feedback: A systematic investigation of college students’ acceptance. *Computer Assisted Language Learning*, *35*(9), 2817-2842. <https://doi.org/10.1111/j.1944-9720.1991.tb00464.x>

**Bionote**

Nguyen Ngoc Duyen is a lecturer and academic specialist at the American Degree Program, Duy Tan University. She holds MAs in TESOL and Public Relations. Her research interests include learner autonomy, foreign language anxiety, second language writing, technology-enhanced language learning, and curriculum design. In 2023, she received the Sheffield Institute of Education Prize for Best Dissertation and the British Council Master’s Dissertation Award 2023-2024, which recognises outstanding UK Master's research with potential impact on the field of English Language Teaching (ELT).

**Appendices**

**Appendix A: Demographic statistics**

**A1. Gender**

|  |  |  |
| --- | --- | --- |
| **Gender** | **Counts** | **Percentage of total** |
| Male | 5 | 41.7% |
| Female | 7 | 58.3% |
| Other | 0 | 0% |
| **Total** | **12** | **100%** |

**A2. Area of study**

|  |  |  |
| --- | --- | --- |
| **Area of study** | **Counts** | **Percentage of total** |
| Business Administration | 3 | 25% |
| Computer Science | 5 | 41.7% |
| Hospitality and Tourism | 4 | 33.3% |
| **Total** | **12** | **100%** |

**Appendix B: Participant consent form**

**Participant Consent Form**

**TITLE OF RESEARCH STUDY:** Human versus Machine: How IELTS Learners Use AI Feedback on IELTS Writing Task 2 Essays and What They Miss

*Please answer the following questions by ticking the response that applies.*

|  |  |  |
| --- | --- | --- |
|  | **YES** | **NO** |
| 1. I have had details of the study explained to me. |  |  |
| 1. My questions about the study have been answered to my satisfaction and I understand that I may ask further questions at any point. |  |  |
| 1. I understand that I am free to withdraw from the study within the time limits, without giving a reason for my withdrawal or declining to answer any particular questions in the study, without any consequences to my future treatment by the researcher. |  |  |
| 1. I agree to provide information to the researchers under the conditions of confidentiality. |  |  |
| 1. I wish to participate in the study. |  |  |
| 1. I consent to the information collected for the purposes of this research study, once anonymised (so that I cannot be identified), to be used for any other research purposes. |  |  |
| 1. I understand that there are no known risks or hazards associated with participating in this study. |  |  |

**Participant’s Signature: \_\_\_\_\_\_\_\_\_ Participant’s Name (Printed): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

A picture containing insect

Description automatically generated**Date: \_\_\_\_\_\_\_\_\_\_\_ Contact details: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Researcher’s Signature: Researcher’s Name (Printed): NGOC DUYEN NGUYEN**

**Researcher's contact details:** 03 Quang Trung, Danang, Vietnam **Phone:** (+84) 0336114937

**Please keep your copy of the consent form and the information sheet together.**

**Appendix C: Feedback form**

**POST-EDITING FEEDBACK FORM AFTER USING AI TOOLS**

Full Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Student ID: \_\_\_\_\_\_\_\_\_\_\_ Date of Editing: \_\_\_\_\_\_\_\_\_\_\_\_\_

**1. Which aspects of the writing did you edit after using ChatGPT? *(Check all that apply)***

|  |  |
| --- | --- |
| ☐ Grammar ☐ Vocabulary ☐ Expression / Paraphrasing ☐ Structure / Logic of the essay ☐ Content ideas | ☐ Relevance to the prompt ☐ Coherence and cohesion of ideas ☐ Other (please specify): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

**2. Briefly describe the main changes you made to your writing (about 3–5 sentences):**..........................................................................................................................................................

**3. How did ChatGPT help you improve your writing?**..........................................................................................................................................................

**4. In your opinion, is the feedback from ChatGPT reliable? Why or why not?**..........................................................................................................................................................

**5. What difficulties did you face when using ChatGPT to edit your writing?**

|  |  |
| --- | --- |
| ☐ Difficult-to-use interface ☐ Did not fully understand ChatGPT’s suggestions ☐ Feedback was not relevant to the prompt | ☐ None ☐ Other (please specify): \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |

**6. Did you use any additional tools/sources besides ChatGPT?**☐ Yes → If yes, please specify: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
☐ No

**7. Any suggestions or further comments on using ChatGPT to improve academic writing skills:**..........................................................................................................................................................

**Appendix D: Interpretation of Kappa values proposed by Landis and Koch Kappa Interpretation (1977)**

|  |  |
| --- | --- |
| **Kappa** | **Interpretation** |
| <0.00 | Poor |
| 0.00-0.20 | Slight |
| 0.21-0.40 | Fair |
| 0.41-0.60 | Moderate |
| 0.61-0.80 | Substantial |
| 0.81-1.00 | Almost perfect |

**Appendix E: Data analysis from feedback form**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Category** | **Student 1** | | **Student 2** | | **Student 3** | | | **Student 4** | **Student 5** | | **Student 6** |
| **Changes made after using ChatGPT** | Grammar, vocabulary, idea development, and improved coherence through transitions | | Grammar, vocabulary, paraphrasing, and smoother phrasing | | Focused mainly on idea development and adding clarification to weak points | | | Made minor grammar edits only (tense, verb forms) | Corrected grammar errors, upgraded vocabulary for academic tone | | Improved vocabulary choices and adjusted transitions for paragraph clarity |
| **Main improvements reported** | Fixed verb tense and sentence structure errors; used more academic words; improved transitions | | Clearer sentences, stronger vocabulary, and more logical introductions and conclusions | | Expanded underdeveloped arguments and improved clarity between points | | | Fixed basic grammar issues like subject–verb agreement | Improved lexical resource with more precise word choices; eliminated redundancy | | Better flow between ideas; refined topic sentences and paragraph unity |
| **Usefulness of ChatGPT feedback** | Helpful for detecting small errors and making writing more fluent and academic | | Very useful in refining both sentence-level and paragraph-level logic | | Helped develop clearer argument structure and avoid repetition | | | Useful for spotting and fixing grammar issues | Helped polish vocabulary, but needed more specific arguments | | Useful for phrasing suggestions and general fluency, but missed deeper logic issues |
| **Trust in ChatGPT feedback** | Somewhat reliable; results varied when using multiple AI tools | | Generally trustworthy but still reviewed against the original task | | Limited trust in deeper content suggestions; mostly relied on AI for sentence-level issues | | | Trusted completely for simple grammar error correction and word choice, improve readability | Trusted for vocabulary and grammar, but cautious about task alignment | | Partially trusted; AI did not always understand the purpose of each paragraph |
| **Difficulties encountered** | Feedback not always aligned with IELTS task demands | | Sometimes unclear why a particular sentence was flagged for revision | | No major difficulties; used feedback selectively | | | No reported issues | AI missed subtle task misinterpretations and gave off-topic suggestions | | Occasionally confusing phrasing or suggestions not relevant to task |
| **Other tools used** | Write&Improve, Lexibot | | Gemini | | None | | | None | Grammarly – inconsistent feedback | | Lexibot – inconsistent feedback |
| **Category** | | **Student 7** | | **Student 8** | | **Student 9** | **Student 10** | | | **Student 11** | **Student 12** |
| **Changes made after using ChatGPT** | | Grammar, paraphrasing, added transition phrases for better cohesion | | Grammar and tone adjustments for a more formal academic style | | Grammar corrections and idea development through expanded examples | Grammar corrections and clearer logical ordering | | | Structural adjustments and use of transitions to improve paragraph flow | Grammar and vocabulary revisions to enhance lexical accuracy |
| **Main improvements reported** | | Reworded unclear sentences; made writing smoother with appropriate linking devices | | Adjusted tone from informal to academic; improved grammar accuracy | | Added examples to support claims; fixed run-on sentences | Improved coherence between ideas; clarified unclear topic sentences | | | Strengthened structure with clearer introductions, topic sentences, and supporting details | Chose more academic vocabulary and refined collocations |
| **Usefulness of ChatGPT feedback** | | Helped fix awkward phrasing and gave better alternatives for transitions | | Helped shift style to academic tone and improved sentence structure | | Useful for grammar and vocabulary, not content | Helped reorganize points logically, but lacked critical evaluation of idea quality | | | Helped correct structural flaws but needed more focus on content development | Supported clearer expression and helped with mechanics |
| **Trust in ChatGPT feedback** | | Somewhat helpful, but needed human validation | | Fully trusted for sentence-level revision | | Used AI feedback only as a reference, not for critical content decisions | Mixed feelings: helpful for fluency, but weak on idea strength | | | Not fully trusted—used for editing only | Found ChatGPT suggestions reliable for surface corrections |
| **Difficulties encountered** | | Minor confusion with ambiguous feedback | | No significant issues | | Concerned about becoming too dependent on AI | Suggestion sometimes not relevant to task requirements | | | Sometimes misread purpose of a paragraph or misidentified topic-related issues | No reported problems |
| **Other tools used** | | None | | None | | Grammarly | None | | | Gemini | None |