AI FOR IELTS WRITING PREPARATION: REVIEWING CHATGPT-GENERATED ANSWER MOCKUPS AND ITS TEXT EVALUATION

Erina Andriani*
Jakarta International University, Bekasi
andrianierina@gmail.com
*correspondence: andrianierina@gmail.com

Abstract
Generative AI (GenAI) tools, such as ChatGPT, are increasingly used as accessible, fast, and consistent alternatives for language proficiency exam preparation, especially for writing tasks. While GenAI tools are seen as efficient in providing feedback, little research has explored their role in generating answer mockups for IELTS writing tasks and evaluating the mockups they create. This study investigated the use of ChatGPT for IELTS General Training Writing Task 1 preparation by addressing these research questions: "How did ChatGPT generate answer mockups for IELTS General Training Writing Task 1?" and "How did ChatGPT evaluate the answer mockups it created?" This research employed qualitative content analysis, examining 15 mockups created for five IELTS writing prompts across score bands 6, 7, and 8. The findings showed that ChatGPT could generate mockups by adhering to IELTS task prompts and criteria in most cases. ChatGPT could also provide detailed evaluations and scores for each writing criterion. However, discrepancies between the generated mockups' scores and the evaluation scores raise questions about the consistency and accuracy of the assessments. The implications of this study are for ChatGPT as a more accessible option for IELTS preparation and as an evaluation of GenAI tools' responses for future development.

Keywords: ChatGPT, General Training Writing Task 1, Generative AI (GenAI) Tools, IELTS, writing mockup

Introduction

The International English Language Testing System (IELTS) serves as a pivotal benchmark for evaluating English language proficiency. As the use of English as a foreign language has only grown stronger each day, English language proficiency tests such as IELTS play a significant role in global education, immigration, and professional contexts (Andriani & Ena, 2022). Whether for academic or professional opportunities, people who want to thrive in an English-speaking environment will first be tested on their language abilities. IELTS gains recognition globally and comprehensively assesses non-native speakers' English

language skills (Sun, 2023), making it one of the most used and versatile English language proficiency tests. Among the skills assessed in the IELTS test, writing is perhaps one of the most challenging (Ariyanti & Anam, 2021), as it requires the test-takers to not only create grammatically correct sentences but also make them into coherent and well-organized texts (Ginting et al., 2023; Lichtinger, 2018). The IELTS writing test is separated for academic and general training purposes, and it consists of Task 1 and Task 2 tests (Sun, 2023). The IELTS test takes coherence and cohesion, vocabulary, grammatical accuracy, and task alignment into account (ielts.org, accessed in August 2024), making it a comprehensive yet challenging test. To successfully achieve the desired IELTS score and eventually have good English proficiency, people need accessible resources to accommodate their language learning.

To assist the language learning and test preparation processes, people have opted for generative AI-powered tools in recent years. Generative AI-powered tools (GenAI) such as ChatGPT, Gemini, and Copilot have increasingly become more popular among language learners (Alexopoulou et al., 2017; Huang et al., 2023; Lin & Chang, 2020). Those tools are user-friendly, effective, and have free access, making them widely used for lots of purposes, including language learning (Assad, 2024; Gayed et al., 2022; Jeanjaroonsri, 2023; Zhao et al., 2024). Some GenAI tools people often use for their learning and improving grammar accuracy include ChatGPT, Grammarly, Gemini, Quillbot, and Wordtune (Athanassopoulos et al., 2023; Ginting et al., 2023; Sutrisno, 2023; Van Horn, 2024; Zhao et al., 2024). GenAI tools can be used as virtual tutors (Barrot, 2023; Mogavi et al., 2024; Lin, 2023) to provide learning programs based on the desired learning goals, create materials, and give feedback. People also use those tools to create learning models they can imitate or learn from, as GenAI tools can provide human-like content (Kohnke et al., 2023; Wang, et al., 2024). Those tools can also evaluate learners' work based on the predetermined criteria they set for their work, increasing their work' quality and eventually their language learning (Koraishi, 2023; Kurniati & Fithriani, 2022; Wilson et al., 2022; Wilson & Roscoe, 2020). GenAI tools can be accessible tools to learn and practice, beneficial for test-takers, such as for IELTS.

Since the emergence of the internet, which supports globalization as a gateway to higher education and career opportunities, the need for English skills and proficiency has increased rapidly. This case leads to the need for comprehensive and globally-trusted tests to assess people's English knowledge. The International English Language Testing System (IELTS) is one of the most well-known tests to check English proficiency (Sun, 2023), established in 1980 (ielts.org, accessed in August 2024). IELTS has gained a great reputation as a reliable English proficiency test, notably influential for university admissions, immigration, professional certifications, and other career opportunities (Sutrisno, 2023). This test has two types, including IELTS academic (for university or higher education admission and professional registration purposes) and IELTS general training (for immigration and work purposes) (Sun, 2023), with the addition of IELTS for UKVI (United Kingdom Visas and Immigration) (ielts.org, accessed in August 2024). This test assesses the reading, writing, speaking, and listening skills of non-native English speakers (Sun, 2023). Among those skills, the writing part is one of the most challenging, especially for those in an EFL environment (Ariyanti & Anam, 2021),

due to the high risk of language errors, improper writing style, and lack of cohesion and coherence (Sun, 2023).

In assessing writing skills, IELTS in Academic and General Training types has two sections of writing tasks, including Task 1 and Task 2. In IELTS Academic Writing, Task 1 is about writing a descriptive report of a data visual representation (graph, chart, table, or diagram), while in General Training, Task 1 is about writing a formal letter. However, both IELTS Academic and General Writing Task 2 are about writing an exposition about a certain topic (Sun, 2023; Sutrisno, 2023). The focus of this study is the IELTS General Training Writing Task 1 about writing a formal letter, and the IELTS website provides the assessment criteria for this task, including task achievement, vocabulary, coherence and cohesion, and grammatical accuracy (ielts.org, accessed in August 2024). Mastering these criteria is crucial and directly impacts the overall band score. Writing skill mastery is essential as it reflects the ability to communicate effectively in academic and professional settings (Emran et al., 2024). Nevertheless, English non-native speakers often face challenges such as organizing the text, using the most appropriate vocabulary, and maintaining grammatical accuracy (Andriani & Ena, 2022; Budjalemba & Listyani, 2020; Maharani & Santosa, 2021). Therefore, to enhance English language proficiency and eventually achieve the best IELTS score, test-takers need focused practice and adequate preparation for the test.

To improve English language proficiency and gain desirable scores on the IELTS test, test-takers must prepare and practice their English language skills and task-specific problems. To effectively prepare for the test, people involve various strategies, such as enrolling in preparation courses, learning practice drills from books, videos, or websites, or observing various task problems and their answers in different score bands (Andriani & Ena, 2022; Triana et al., 2023). With the growth of Generative AI tools and their popularity, people have started to employ GenAI tools to enhance their preparation. GenAI tools for IELTS preparation offer advantages, including free access, personalized learning experiences, adjustable learning goals and styles, and the freedom to set task criteria to achieve the desired score (Koraishi, 2023; Wang et al., 2024). By providing tailored feedback and targeted practice exercises, AI can help learners improve specific skills (Athanassopoulos et al., 2023; Emran et al., 2024; Van Horn, 2023; Wale & Kassahun, 2024; Xiao & Zhi, 2023; Zhang, 2024). This personalized approach not only makes study sessions more productive but also boosts confidence, ultimately leading to better performance on the test.

From its viral beginning until now, GenAI tools have kept emerging and continued to evolve, rapidly entering many aspects of life, including education. Its integration into educational settings has sparked significant interest and debate, reshaping traditional learning methodologies and offering new and broader ways to learn. Some types of GenAI tools that can be used in language learning include speech technology, machine translation, chatbots, AI-generated content, and virtual assistants. Some GenAI tools applicable for language learning include Google Translate, Alexa, Quizizz AI, ChatGPT, Grammarly, Gemini, Quillbot, and Wordtune (Athanassopoulos et al., 2023; Ginting et al., 2023; Sutrisno, 2023; Van Horn, 2024; Zhao et al., 2024). A study by Lavidas et al. (2024) showed the perceived benefits, such as enjoyment, habitual, and performance aspects, were the reasons for students in using GenAI tools. For writing skills, GenAI tools can help

in various aspects such as academic, business, and creative writing (Alberth, 2023; Assad, 2024; Ginting et al., 2023; Imran & Almusharraf, 2023; Maghamil & Sieras, 2024; Song & Song, 2023; Utami et al., 2023; Wafa et al., 2024; Zhou & Zhang, 2024). Language learners can use GenAI tools to outline, revise, proofread, and reflecting (Su et al., 2023), be personal tutors to help organize the writings logically and coherently (Marzuki et al., 2023), and provide the answer samples to actually see the aspects they need to improve in their own writing (Koraishi, 2023). Though greatly beneficial, many studies highlighted concerns about the use of GenAI tools for language learning, such as critical thinking, academic dishonesty, reduced self-efficacy, accuracy, relevance, and also technical issues to formulate the best command to get the best-fitting results (Assad, 2024; Van Horn, 2024; Zhang, 2024). Despite the challenges, GenAI tools can indeed help in the learning process (Wardat et al., 2023), and thus are beneficial for IELTS test preparation.

One of the GenAI tools that can help with IELTS test preparation is ChatGPT. Chat Generative Pre-Trained Transformer (ChatGPT) is an AI-powered chatbot that is equipped with a massive volume of knowledge to provide users with information in a human-like delivery (Barrot, 2023). Its strengths are to provide human-like content and filter out unrequired information beyond what was requested (Hong, 2023; Yan, 2023). They let users save time browsing and screening information, and focus on just the needed information. In addition, ChatGPT is interactive and can give instant answers, making language learners more engaged with real-time feedback (Rahman et al., 2023; Utami et al., 2023; Xiao & Zhi, 2023). Despite the merits, ChatGPT still has its limitations. Barrot (2023) criticized its sensitivity to minor adjustments in the command wordings and the tendency to follow specific structures. This can be a minus for IELTS writing preparation, as users need to be wary about the wording and also its repetition (Buriak et al., 2023). This tool also cannot browse the internet (openai.com; accessed in August 2024), and while it is equipped with vast information, it may not be up-to-date. ChatGPT itself warns users to verify the reliability and accuracy of the information it provides (Le & Mohd, 2024; openai.com; accessed in August 2024). Many studies, however, highlight the profound benefits of ChatGPT use for IELTS test preparation (Rahman et al., 2023; Le & Mohd, 2024; Song & Song, 2023; Sutrisno, 2023; Wang, et al., 2024).

A way ChatGPT accommodates IELTS writing preparation is by giving writing mockups for learners to observe and learn from. Having writing mockups for IELTS preparation can help provide realistic illustrations, immediate feedback, and adaptation to the test conditions (Meniado, 2023; Sutrisno, 2023). One significant advantage highlighted by previous studies is ChatGPT's ability to deliver personalized feedback based on individual writing samples, which helps learners address specific areas for improvement efficiently (Barrot, 2023; Mogavi et al., 2024; Wang, et al., 2024). Moreover, ChatGPT can offer diverse practice questions and model responses, enhancing the overall preparation experience. However, ChatGPT has several limitations. Studies suggest that while ChatGPT can simulate conversation and provide feedback, it may not fully grasp the nuanced context and creativity required for high-quality writing (Vicente-Yagüe-Jara et al., 2023). Additionally, excessive dependence on AI tools like ChatGPT could potentially impede the development of critical thinking and original writing skills that are crucial for the IELTS test (Thorp, 2023). Thus, while ChatGPT provides valuable

support, it is essential to integrate its use with other study methods to ensure comprehensive preparation.

Aside from using ChatGPT to provide writing mockups, learners can also use ChatGPT to evaluate their own writing based on IELTS scoring criteria. After having ChatGPT provide the writing mockups based on IELTS problems for learners to observe and learn from, learners can then practice writing their own answers. The next natural step would be assessing their writing based on the IELTS criteria (Koraishi, 2023; Sun, 2023). Evaluating and seeking feedback on their writing is another study method for comprehensive preparation. Learners can employ this method using ChatGPT as their personal tutor (Hong, 2023; Mogavi et al., 2024; Xiao & Zhi, 2023). To ask ChatGPT to help evaluate their writings, learners can prompt their command by including the list of IELTS writing criteria in addition to the request to evaluate their writings. The learners need to clearly prompt their commands in ChatGPT, as it can be sensitive to the wording (Barrot, 2023). The IELTS General Training Task 1 criteria include coherence and cohesion, task achievement, grammatical accuracy, and lexical resource (Sun, 2023; Sutrisno, 2023; Wale & Kassahun, 2024), as provided on the IELTS website (ielts.org, accessed in August 2024). Learners can also specify the score band they aim to achieve, which allows for more personalized feedback tailored to their needs. Using ChatGPT as a tool to assess writing can be advantageous due to its real-time responses, making it more efficient (Rahman et al., 2023; Hong, 2023; Koraishi, 2023). This instant feedback supports interactive learning for students to directly learn ways to enhance their writing in a context-specific setting (Kurniati & Fithriani, 2022; Marzuki et al., 2023; Xiao & Zhi, 2023). It is also arguably consistent, as it does not have bias and subjectivity aspects (Sun, 2023). However, some studies show limitations, such as relevance, whether the responses provided by the chatbot are relevant to the question/prompt the learners give (Assad, 2024). ChatGPT may lack contextual understanding and rely heavily on the prompted criteria (Sun, 2023). This technical issue can lead to inaccuracy in the ChatGPT responses (Barrot, 2023; Van Horn, 2024).

On the topic of ChatGPT use in IELTS preparation, some previous studies shed light on its possible applications, benefits, and challenges. Sun (2023) shared the potential of using ChatGPT as a tool for IELTS writing assessment over manual assessment. The study compared the advantages and challenges of ChatGPT-based and manual assessments. Sutrisno (2023) reviewed the use of ChatGPT to provide writing samples and feedback on the learners' writing for IELTS Writing Task 2. Rahman et al. (2023) conducted a case study on the emergence of GenAI tools and their impact on the English proficiency tests' writing section. Their findings show the ways GenAI tools help English proficiency test preparation, the instances where GenAI tools can be highly beneficial for non-native English speakers' learning process, the drawbacks of using GenAI tools, and the reasons why GenAI tools will not eliminate the writing section in English proficiency tests. Wang et al. (2024) highlighted the potential uses of GenAI tools such as ChatGPT for language teaching, which are applicable in IELTS test preparation. Though previous studies showed the use of ChatGPT in IELTS preparation, they have yet to review its use for IELTS General Training Writing Task 1 preparation. Hence, this study intends to dive into this topic.

Due to the advantages in assisting the language learning process, GenAI tools have become widely used for IELTS test preparation, especially for writing tasks. GenAI tools are very accessible to assist language learners' study and practice with several advantages including instant and personalized feedback, sophisticated human-like content for task examples, interactive assistance, and specified learning objectives (Athanassopoulos et al., 2023; Gayed et al., 2022; Kohnke et al., 2023; Sutrisno, 2023; Wang, et al., 2024). Specifically in writing, GenAI tools serve various purposes, such as brainstorming, translating ideas into texts, outlining, proofreading, revising work, and giving writing strategies (Su et al., 2023; Wang, et al., 2024). Since the emergence of GenAI tools, many researchers have studied the use of AI for language learning and writing (Rahman, et al., 2023; Susanto, 2023; Sun, 2023; Sutrisno, 2023). However, they rarely highlight the use of GenAI tools to provide answer mockups for IELTS writing test preparation and how the GenAI tools themselves evaluate those texts. Therefore, this research aims to study the ChatGPT-generated IELTS General Training Writing Task 1 answer mockups and its text evaluation, as it assesses the effectiveness and reliability of GenAI tools in simulating human-level writing proficiency, potentially revolutionizing educational and testing practices by providing consistent, scalable feedback. The research questions of this study are "How did ChatGPT generate answer mockups for IELTS general training writing task 1?" and "How did ChatGPT evaluate the answer mockups it created?" The implications of this study are for test-takers to have more accessible options for IELTS test preparation and as an evaluation of GenAI tools' writing creations for future development.

Method

This paper explored ChatGPT as a GenAI tool to assist non-native language learners in preparing for their IELTS test. This research used qualitative research with content analysis, as this research aimed to study the text content of ChatGPT-generated IELTS General Training Writing Task 1 mockups and its evaluation of those mockups (Silva, 2017). Five IELTS writing prompts were used: two prompts from IELTS preparation resources, provided on the IELTS website, and three prompts made by ChatGPT. From each of the five prompts, three answer mockups were collected, corresponding to score bands 6, 7, and 8, resulting in a total of 15 mockups. Those 15 samples allowed the researcher to analyze the writing composition in different score bands across several writing tasks while still allowing for more in-depth qualitative analysis. This number also improves on previous studies that used fewer samples (Sutrisno, 2023). This study employed purposive sampling (Creswell, 2012) as the researcher predetermined the criteria for the samples (Etikan & Bala, 2017), including the writing prompts and the score bands of the answer mockups.

To gather the data, the researcher used an observation sheet for both research questions based on the writing task criteria as provided on the IELTS website, including task achievement, coherence and cohesion, vocabulary, and grammatical accuracy (ielts.org, accessed in August 2024). The researcher first set the conversation in ChatGPT by having it define and give the IELTS criteria and score bands. Once ChatGPT was set in the IELTS General Training Task 1 environment, the researcher commanded ChatGPT to generate answer mockups based on the prompts and score bands. For the mockups using ChatGPT's prompts, the

researcher asked ChatGPT to make the prompts first before the answer mockups. The researcher then recorded the answers.

For the second research question, the researcher sent each answer mockup for ChatGPT to assess based on the criteria and score bands. The review on ChatGPT's assessment included these questions: "What score band do you rate this writing mockup to?", "What is your feedback on this text?", and "What needs to be improved in this writing to gain a higher score band?". The first question was given to check ChatGPT's understanding of the IELTS score bands and if the score band it gave was the same as the score band the researcher commanded for the first research question. The second and third questions were given to further see its assessment and provide feedback for learners as a prospective virtual tutor.

The command prompts were adapted from Sun (2023) and Sutrisno (2023) and piloted prior to the main study to ensure clarity and relevance. First, the researcher requested ChatGPT's understanding of IELTS Writing Test 1 before giving the adapted command prompts. In some instances, the command prompts were fine-tuned by requiring more details when it only provided incomplete answers (e.g., giving brief criteria points and 0-9 score bands without further details). To avoid any bias from ChatGPT in its own answer mockups, the researcher stayed logged out of the ChatGPT account and hard-refreshed the page every time the researcher needed ChatGPT to evaluate each mockup. However, the researcher still conditioned ChatGPT with IELTS General Training Task 1 criteria and score bands every time first, before asking for its assessment of each mockup. The researcher then recorded the answer based on the structured observation sheet.

To analyze the data, the researcher evaluated and marked the parts of the mockups based on the IELTS Writing Task 1 criteria and recorded the occurrences of each criterion. The researcher analyzed the mockups by evaluating them against the scoring criteria. Each mockup was examined to see whether it addressed all parts of the writing prompt and presented supporting details (Task Achievement); whether the ideas were organized logically and connected using appropriate discourse markers (Coherence and Cohesion); whether the vocabulary was topic-appropriate, natural, and avoided unnecessary repetition (Lexical Resource); and whether the sentences showed grammatical variety with minimal errors (Grammatical Range and Accuracy). The acquired themes were then categorized into each scoring criterion. The researcher did a similar analysis process on the second research question to check ChatGPT's answer mockup evaluation based on the IELTS Writing Task 1 criteria. The results were then discussed descriptively.

Findings and Discussion

This study investigated how ChatGPT, as a GenAI tool, generated answer mockups for IELTS General Training Writing Task 1 within specific score bands and evaluated these answer mockups. This is to see if ChatGPT can be a reliable alternative to provide model texts for non-native people to learn from, and assess the consistency of its evaluation on the score bands for each mockup. At first, the researcher conditioned the conversation to be about IELTS General Training Writing Task 1 for each session by asking about its knowledge of the said test and the scoring criteria. Then, the researcher gave ChatGPT a command to create answer mockups and collected its answers, before asking it to evaluate those mockups in separate sessions.

ChatGPT-generated mockups for IELTS General Training Writing Task 1

The researcher first set the context in ChatGPT to talk about IELTS General Training Writing Task 1. When the researcher first conditioned the conversation in ChatGPT to be about IELTS General Training Writing Task 1, it could provide the general idea of the test as well as the scoring criteria, as presented in Figure 1. The response not only mentioned the criteria but also provided a brief explanation about each criterion. ChatGPT was then presented with two writing prompts taken from the IELTS website (ielts.org, accessed in August 2024). The researcher also asked ChatGPT to create three more writing prompts. It could create the writing prompts and provide the answer mockups for all prompts for score bands 6, 7, and 8 at once in one response.
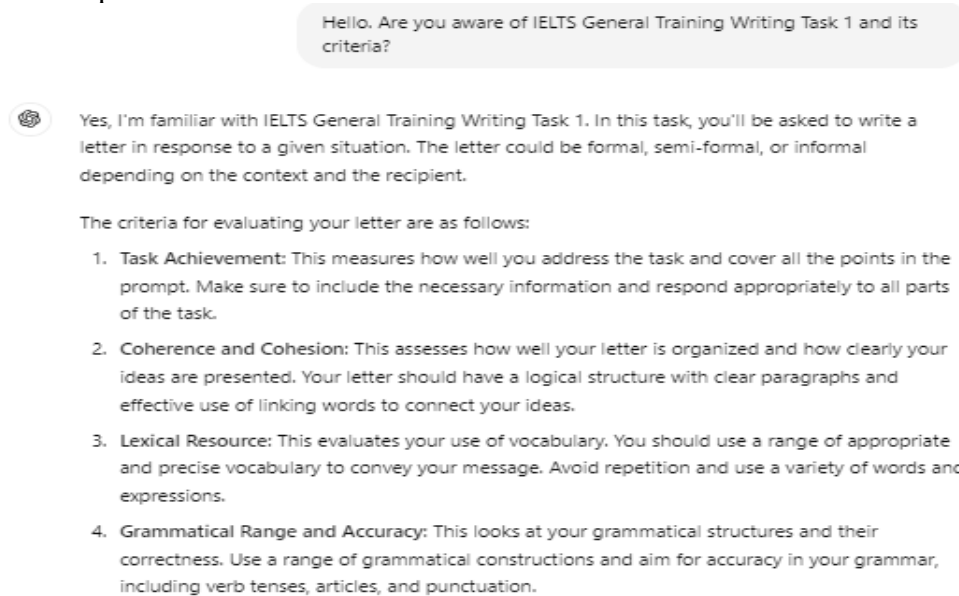

Figure 1. Setting Context in ChatGPT Conversation

In all initiated sessions, ChatGPT could provide the answer mockups based on the prompts for band scores 6, 7, and 8 directly in one response. All answer mockups varied in their lengths or word counts. Even though the mockup length does not necessarily reflect their quality, despite the inconsistencies in length between mockups in bands 6 and 7, the mockups from band 8 were consistently the longest, as shown in Table 1. In addition, all of the ideas to answer the prompts were mainly the same for all score bands, but each band varied the phrasing to mention the ideas and the details to support them. Another interesting finding is that all answer mockups always end with the phrase "Yours sincerely," despite the score bands and prompts.

Table 1. Answer Mockup Word Count

Prompt	Band 6	Band 7	Band 8
Prompt 1 (IELTS1)	198	196	204
Prompt 2 (IELTS2)	141	141	151
Prompt 3 (ChatGPT1)	149	137	164

| Prompt 3 (ChatGPT2) | 122 | 130 | 145 |
| Prompt 3 (ChatGPT3) | 115 | 115 | 156 |

Per the writing criteria, answer mockups made by ChatGPT addressed the task points presented in the writing prompts to satisfy the task achievement. In answering the task points, the mockups presented related information details. The information and ideas to answer the task points were mainly the same for the mockups in each score band, as presented in Examples [1] to [3].

[1] Band 6 prompt 5: Overall, **I found the workshop to be informative and the topic was very interesting**.

However, there were some issues. The workshop **started late**, which shortened the time for the main presentation. Additionally, the **materials provided were not very clear**, and **some of the handouts had errors**.

To improve the workshop, I would suggest **starting on time** and ensuring that **all materials are reviewed and corrected** before distribution. Also, it might be helpful to provide more **interactive elements or practical exercises** to engage the participants better. Thank you for considering my suggestions.

Yours sincerely

[2] Band 7 prompt 5: **The workshop was quite engaging, and I appreciated the depth of the information presented.**

However, there were a few problems. The workshop **began later** than scheduled, which cut into the time allocated for discussions. Additionally, some of the **materials provided were confusing**, with several **errors in the handouts**.

For future workshops, I recommend **starting on time** and **double-checking all materials** for accuracy before distribution. Including more **interactive activities or practical exercises** could also enhance the experience for participants.

Thank you for taking my feedback into account.

Yours sincerely

[3] Band 8 prompt 5: I found the **workshop to be very insightful and well-organized, with a well chosen topic that was relevant and engaging.**

Nonetheless, there were a few issues that impacted the overall experience. The workshop **began later** than planned, which unfortunately reduced the time available for the main content and discussions. Additionally, the **materials provided had some errors** and **were not as clear** as they could have been, which affected their usefulness.

To enhance the quality of future workshops, I suggest ensuring that the workshop **starts on schedule** and thoroughly **reviewing all materials** for accuracy before distribution. Additionally, incorporating more **interactive components or hands-on exercises** could greatly benefit participants and make the sessions more dynamic.

Thank you very much for considering these suggestions. I look forward to attending future events.

Yours sincerely

The task points from the writing prompt for those mockups are as follows (source: ielts.org):
- *Describe what you thought about the workshop.*
- *Explain any problems you encountered during the workshop.*
- *Suggest ways to improve the workshop for future participants.*

Though the mockups for each score band worded the details differently, all of them had the same details. Examples [1] to [3] show answer mockups from the fifth prompt for bands 6 to 8, respectively. To answer the first task point, all mockups from bands 6 to 8 shared positive responses about the workshop, highlighting the informative aspect. For the second task point, the problems included that the workshop started late, and the problems with the materials, such as errors and a lack of clarity. For the last task, all mockups mentioned starting on time, reviewing materials, and including interactive and practical activities as suggestions for future improvement. It might be ChatGPT's intention to illustrate the comparison for different score bands in presenting the same details to answer the task points. However, by presenting the same details, it may seem repetitive. This is in contrast to Sutrisno (2023) that ChatGPT can provide different samples and not be repetitive. Showing comparisons of each mockup can be beneficial, but if the users intend to have variety in the mockups, they may need to specify detailed variety in the command to avoid the answer details being similar. On a related term, even though the closing greeting was not regulated in the prompts, all fifteen mockups used 'Yours sincerely' as the closing remark, as presented in Examples [1] to [3].

As requested in the commands, the answer mockups generated by ChatGPT followed the presented writing prompts. All answer mockups strictly followed the things to write and not to write as stated in the prompts, except for the second prompt, where there was an instruction to write at least 150 words, but only the band 8 mockup fulfilled this condition, as presented in Examples [4] to [6].

[4] Band 6 prompt 2: Dear Sir or Madam,

I am writing to share my thoughts on the local public library and suggest some improvements. I generally appreciate the library's extensive collection of books and the quiet environment, which makes it a great place to study.

However, there are a few aspects I find problematic. The library's computer facilities are often out of order, and it can be frustrating to wait for a computer when needed. Also, the seating areas are not very comfortable, and there are not enough tables for group study sessions.

To improve the library, I suggest that you invest in upgrading the computer facilities and ensure they are well-maintained. Additionally, adding more comfortable seating and more tables for group work would make the library more user-friendly and supportive of different study needs.

Thank you for considering my suggestions.

Yours sincerely,

[Your Name]

[5] Band 7 prompt 2: Dear Sir or Madam,

I am writing to provide feedback on the local public library and to offer some suggestions for its improvement. I value the library for its

comprehensive range of books and the serene atmosphere, which is ideal for focused study and research.

Despite these positives, there are a few issues that need addressing. The library's computer stations are frequently down, which disrupts users who need to access digital resources. Additionally, the seating options are limited and not very ergonomic, which can be uncomfortable for extended periods of reading or studying.

I recommend that the library upgrade its computer facilities to ensure they are consistently operational. Improving the seating arrangements by providing more comfortable chairs and additional tables for group work would also greatly enhance the user experience.

I hope these suggestions will be helpful.

Yours sincerely,

[Your Name]

[6] Band 8 prompt 2: Dear Sir or Madam,

I am writing to offer feedback on the local public library and to suggest some potential improvements. I have always enjoyed the library's vast collection of books and the calm, conducive environment it provides for studying and reading.

However, there are some areas that could be improved. The computer facilities are often unreliable, leading to frustration among users who need access to online resources. Additionally, the current seating arrangements are inadequate, with limited comfortable seating and insufficient tables for collaborative study.

To enhance the library's services, I propose investing in modernizing the computer facilities to ensure they are always functional. Furthermore, expanding the seating options to include more ergonomic chairs and additional tables for group study would greatly improve the overall user experience and support various study needs more effectively. Thank you for considering my suggestions. I look forward to seeing these improvements.

Yours sincerely,

[Your Name]

The task points for those mockups are presented in Figure 2.

Knowing the criteria and score bands, please make writing mockups of
this prompt, for score bands 6, 7, and 8.

Your local public library wants to make improvements to their services
and facilities. In order to get ideas from the public, they have asked
library users to send them suggestions in writing.
Write a letter to the librarian. In your letter
• describe what you like about the library
• say what you don't like
• make suggestions for improvements
Write at least 150 words.
You do NOT need to write any addresses.
Begin your letter as follows:
Dear Sir or Madam,

Figure 2. Second Writing Prompt (source: ielts.org)

All mockups fulfilled the given instructions from the prompts, but each score band presented the level of detail in the ideas differently. Perhaps due to the fewer details in bands 6 and 7, the given mockups did not follow the prompt's request to write at least 150 words. The mockups for bands 6 and 7 had 141 words, while band 8 had 151 words. This finding contradicts the finding from Sutrisno (2023) that ChatGPT can provide answer samples based on the given prompts. However, as other mockups in this study could satisfy the tasks in every other writing prompt and only this prompt used a word limit, the researcher deems it a special case and perceives that further research is needed to generate a more representative conclusion.

As for grammar accuracy, all mockups were free from major grammatical errors, though some can be rephrased to increase clarity. An example is presented in Example [7].

[7] Band 8 prompt 5: …Furthermore, expanding the seating options to **include more ergonomic chairs** and additional tables for group study would greatly improve the overall user experience and support various study needs more effectively.…

Grammatical accuracy is another crucial aspect of English proficiency assessed in IELTS. The answer mockups presented by ChatGPT had excellent grammatical accuracy, as the responses given by GenAI tools are often grammatically accurate (Sun, 2023; Sutrisno, 2023). However, grammatical accuracy goes beyond the syntactical structure. For example [7], using more ergonomic chairs phrase may cause ambiguity due to a lack of clarity on whether it means the chairs should be more ergonomic or increase the number of ergonomic chairs. This structural ambiguity may affect the grammatical accuracy point, and it can be avoided by rearranging the structure or putting indicators (Kurniasari, 2017). GenAI tools are increasingly used to ensure the grammatical accuracy of one's writing (Fitria, 2021) and ChatGPT is known to provide grammatically accurate responses (Maghamil & Sieras, 2024). The findings of this study mostly concur with the previous studies,

though findings like in Example [7] should be taken into account when using GenAI tools to always check the contents provided by GenAI tools.

The coherence and cohesion points varied across all mockups. Some mockups from the same score band had arguably better use of coherence and cohesion than others. Some examples were presented in Examples [8] and [9].

[8] Band 6 prompt 1: I am writing to bring to your attention several issues I am experiencing with my current accommodation. I share a room in college with another student, and there are multiple problems that make it very hard for me to focus on my studies.

Firstly, the main issue is the lack of privacy. The room is quite small, and my roommate's frequent visitors create a lot of noise, which is very distracting. **Additionally**, my roommate often studies late into the night with the light on, which interferes with my sleep and makes it difficult for me to concentrate the next day.

Due to these problems, I am finding it very challenging to manage my coursework effectively. The constant interruptions and lack of quiet time are significantly affecting my productivity and academic performance.

I would prefer to move to a single room where I can have my own space and a quieter environment. **This would greatly help** me in managing my studies more efficiently and maintaining a better balance between my academic and personal life.

[9] Band 6 prompt 4: I am writing to inquire about staying at your bed and breakfast. My family and I are planning a vacation from June 10th to June 15th. There will be four people in our group.

We would like to have a room with two double beds if possible. **Also**, one of us has a dietary restriction and needs gluten-free options for breakfast. If you have any recommendations for nearby activities or places to visit, **that would be very helpful as well**.

Could you please let me know if you have availability for these dates and what the cost would be for the stay? I would appreciate a prompt response so we can make arrangements.

Examples [8] and [9] were both from score band 6. Despite being in the same score band, Example [8] used more transitions to move from one idea to another. In Example [8], when moving the idea from the context in the first paragraph to the details in the second paragraph, the text used a transition device *'firstly'*. Though it was not followed by secondly, to continue to a second problem, the text used *'additionally'*. Moving to the third paragraph, to indicate that the problem details were finished and change the topic to the impact, the text used *'due to these problems'*. It could use another transition device to start the fourth paragraph, but this mockup used various transition devices to connect ideas. In contrast, example [9] showed less use of transition devices to connect ideas, resulting in poorer coherence and cohesion compared to Example [8], despite being in the same score band. This inconsistency might be due to the different prompts used, and Zhou and Zhang (2024) also mentioned that ChatGPT responses were inconsistent based on the posed context. Therefore, it is important to have the human aspect in addition to the GenAI tools in generating and evaluating responses (Rahman et al., 2023; Sun, 2023).

The last criterion for IELTS Writing Task 1 is lexical resource. Regarding this criterion, all provided mockups used varied vocabulary that fit the context and avoided repetition. Though the mockups often used synonyms to demonstrate diverse expressions and clarity within the same context, sometimes the mockups used the same base word in different word classes instead of using the synonym, as presented in Example [10].

[10] Band 8 prompt 2: I am writing to offer feedback on the local public library and to suggest some potential **improvements**. I have always enjoyed the library's vast collection of books and the calm, conducive environment it provides for studying and reading.

However, there are some areas that could be **improved**. The computer facilities are often unreliable, leading to frustration among users who need access to online resources. Additionally, the current seating arrangements are inadequate, with limited comfortable seating and insufficient tables for collaborative study.

To enhance the library's services, I propose investing in modernizing the computer facilities to ensure they are always functional. Furthermore, expanding the seating options to include more ergonomic chairs and additional tables for group study would greatly **improve** the overall user experience and support various study needs more effectively.

Thank you for considering my suggestions. I look forward to seeing these **improvements**.

Example [10] used different varieties of the base word 'improve'. This word appeared once in each paragraph, though in different forms (*improvements, improved, improve*). This is somewhat of a strategy to avoid repeating the same exact word. Though technically not repeated, using the synonyms of the said word would accentuate wider lexical resources, especially remembering its score band. The benefits of using ChatGPT for learning wide vocabulary usage are aligned with Imran and Almusharraf (2023) and Song and Song (2023), though sometimes people still need to check the writing content to avoid special cases, as in the Example [10].

In addition to the answer mockups, the researcher also took an interest in the prompts that ChatGPT created. Aside from learning from the answer mockups, learners can also practice their writing by finding IELTS-like writing prompts they can write about. The prompts themselves can also affect how the answer mockups turn out. Among the five writing prompts used in the research, three of them are created by ChatGPT. The ChatGPT-created prompts were similar to the prompts provided on the IELTS website regarding the exclusion of the address in the letter. However, there are some differences, as shown in Examples [11] and [12].

[11] You live in a room in college which you share with another student. **However, there are many problems with this arrangement and you find it very difficult to work.** Write a letter to the accommodation officer at the college. In the letter
• describe the situation
• explain your problems and why it is difficult to work
• say what kind of accommodation you would prefer
Write at least 150 words. You do NOT need to write any addresses. Begin your letter as follows: **Dear Sir or Madam,**

[12] You have recently moved to a new city for work and are staying in a rented apartment. Write a letter to your landlord. In your letter:
• Describe your experience with the apartment so far.
• Explain any issues you have encountered.
• Request any necessary repairs or improvements.
You do NOT need to write any addresses. Begin your letter as follows:
Dear [Landlord's Name],

Example [11] is the prompt provided by the IELTS website, while example [12] is the ChatGPT-generated prompt. Those two prompts showed some differences in directing the writing content. The first one is about how to begin the letter. The IELTS prompts usually ask to use 'Dear Sir/Madam', as in Example [11], but in the ChatGPT-made prompts, they asked to begin with names, for example: 'Dear [Landlord's Name]', as in Example [12]. The other difference is that the prompts from ChatGPT lacked the context at the beginning as to why the test-takers should write the letter. For example, the prompt from the IELTS website stated that the letter is due to the arrangement issues with the roommate, causing the writer to have difficulty working, as presented in Example [11]. However, in the prompt made by ChatGPT, the beginning only set the context and then the things to write, not clearly stating the purpose of the letter. The test-takers need to read the whole prompt to roughly configure the purpose of the letter. This finding is aligned with Sutrisno (2023) that ChatGPT can be used to generate IELTS writing prompts on which learners can base their writing. It could be an alternative to provide writing prompts, so learners can have more chances to practice writing using a wider variety of prompts. Even though there were some slight differences between the IELTS-provided prompts and ChatGPT-generated ones, the instructions were clear and still doable to be a writing basis.

Upon researching the IELTS General Training Writing Task 1 mockup generation by ChatGPT, it was capable of generating writing mockups that align well with the IELTS Writing Task 1 criteria, producing responses in score bands 6, 7, and 8 based on the given prompts. The majority of these mockups adhere to scoring criteria. Additionally, ChatGPT could generate alternative writing prompts that, while not identical to traditional IELTS prompts, offered variety for practice. However, there were instances where some mockups did not fully comply with the prompts or criteria, highlighting the importance of human oversight in evaluating the generated content. This suggests that while ChatGPT can serve as a useful tool for generating writing mockups, human involvement is still necessary to ensure the accuracy and quality of the responses.

ChatGPT's evaluation on the answer mockups for IELTS General Training Writing Task 1

To collect the data for the second research question, the researcher first conditioned the conversation in ChatGPT to talk about IELTS General Training Writing Task 1, especially the scoring criteria and bands. An interesting finding was immediately apparent in this step. During the fifteen attempts, ChatGPT's responses mostly directly provided the criteria and bands, as presented in Figure 3.

[13] You said: Are you aware of IELTS General Training Writing Task 1 criteria and score bands?

ChatGPT said: Yes, I'm familiar with the IELTS General Training Writing Task 1 criteria and score bands. Are you looking for details on a specific aspect or need help with a particular task?

Figure 3. Common Response Template

The majority of the responses mostly responded to the conditioning question by providing a brief explanation of the writing criteria and score band. However, in three instances, ChatGPT did not directly explain the scoring criteria and bands, as presented in Example [13]. Instead of directly explaining, it unassumingly confirmed its familiarity with IELTS writing criteria and score band, answering the yes/no question, and gave a probing question for further details. This finding is supported by Shaib et al. (2024) that ChatGPT responses often use the same template and organization, as the responses mostly follow the pre-trained data, though exceptions in the organization format might happen.

After conditioning the room chat, the researcher asked ChatGPT to evaluate the answer mockups it had made in other sessions. In all conversation attempts to ask ChatGPT to assess writing mockups, ChatGPT could provide its evaluation and the score band it viewed the mockups to be in. In delivering the score band, ChatGPT gave a score for each criterion and gave the overall estimated score at the end. Sometimes, the scores for all criteria are the same, as in Example 14. However, sometimes it gave different scores for different criteria, and then the overall estimated score was taken from the average of all scores, as in Example 15.

[14] Band 8 prompt 1: Task Achievement: You address the task well by…
Score Band Estimate: 7
Coherence and Cohesion: Your letter is organized logically and flows smoothly. You have… Score Band Estimate: 7

Lexical Resource: Your vocabulary is appropriate and varied, but there's some room for improvement… Score Band Estimate: 7

Grammatical Range and Accuracy: Your grammar is accurate, with correct sentence structures and punctuation… Score Band Estimate: 7

[15] Band 7 prompt 1: Task Achievement (Band 7):…

Coherence and Cohesion (Band 8):…

Lexical Resource (Band 7):…

Grammatical Range and Accuracy (Band 8):…

Overall Band Score: 7.5

In addition to giving some feedback, ChatGPT assessed the mockups by giving scores for each criterion. The scores were based on the score bands to portray the achievement of each criterion accordingly. In Example [14], all criteria got the same score band, resulting in the overall band being 7. However, sometimes not all aspects were achieved equally. Therefore, ChatGPT could give different scores for each aspect to indicate the stronger and weaker points of the writing, as in Example [15]. There, it gave band 8 to coherence and cohesion, and also grammatical accuracy, indicating the stronger points of the writing. Meanwhile, the other criteria got a band 7. This is in line with the findings of Sutrisno (2023) that ChatGPT could provide scores for each IELTS criterion, and they could be different from one another.

Looking back at Examples [14] and [15], the given overall scores were different from the mockups' "actual" scores that the researcher requested for the first research question. In fact, among 15 mockups, only six were given the same score as the initially requested score from the first research question. Among the nine different scorings, five mockups were rated higher and four mockups were rated lower, as presented in Figure 4.
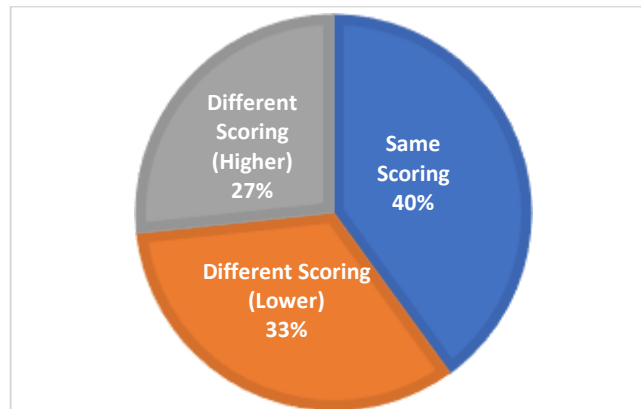
Figure 4. ChatGPT's Estimated vs. 'Actual' Band Scores

Most of the mockups that were scored lower were from band 8. All band 8 mockups were rated lower into band 7, except for the mockup from prompt 4, as presented in Figure 5. Another sample rated lower was from band 7, given a score of 6. Similarly, the mockups that were rated higher were mostly from band 6, except the mockups from prompts 1 and 5. Another sample being rated higher was from band 7, given the score of 7.5, presented in Example [15].
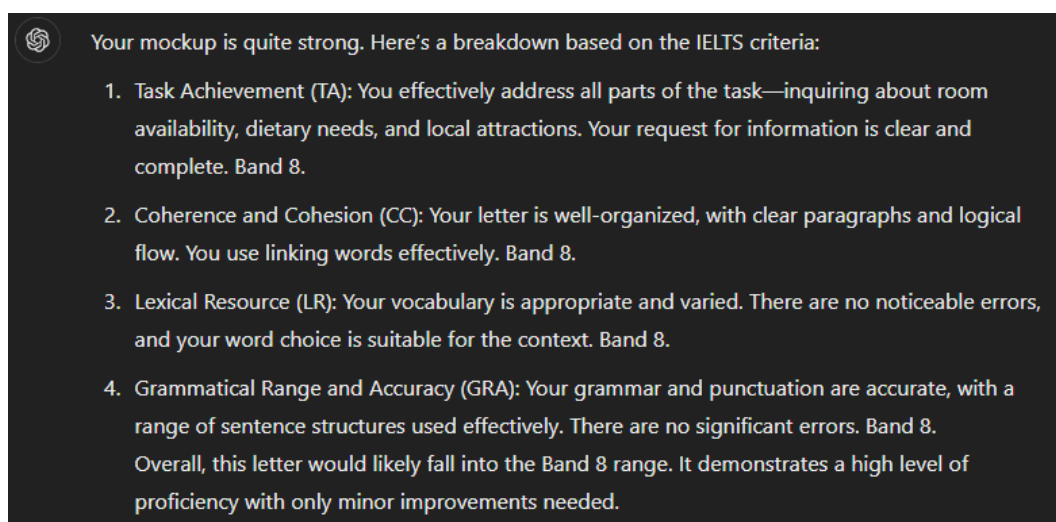
Figure 5. Aligned Scoring from the Generation Stage and the Evaluation Stage

The mockup being evaluated in Figure 5 is from band 8, and ChatGPT could give the same score for this mockup. Still, only 40% of the samples were scored according to the requested score that the researcher commanded in the generation stage. This finding challenges the finding by Sun (2023) about the consistency of GenAI tools to evaluate IELTS writing, which showed the strength of GenAI tools over human assessors. This inconsistency sparks confusion about whether the mockup's scoring or the evaluation's scoring is the most accurate band.

In giving its scoring on the mockup, ChatGPT briefly mentioned things that the mockups could improve. When asked about its feedback and things to improve, ChatGPT gave more detailed responses, stating the strengths of the mockups and how to improve them, as in Figure 6.

Figure 6. ChatGPT's Feedback and Improvement Common Format

However, in the last 3 conversations, ChatGPT did not tell the mockup's strengths in its feedback and directly gave ways to improve the letters, as in Figure 7.

> What is your feedback on this text? And what needs to be improved from this writing to gain a higher score band?
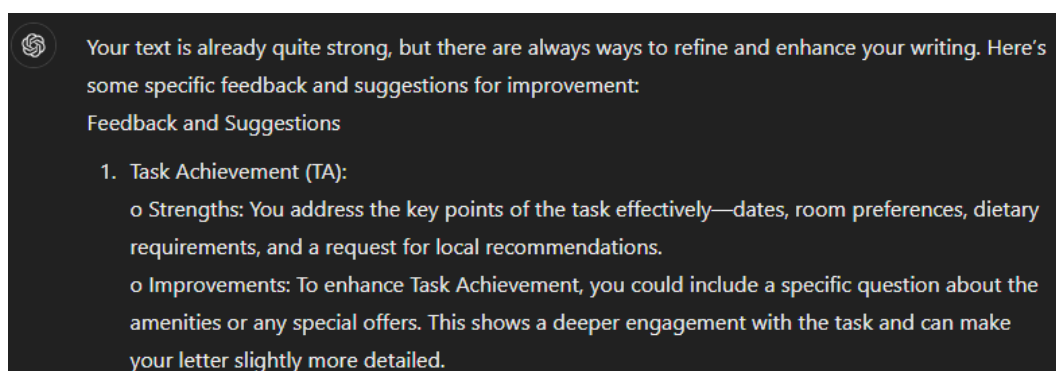
Your text is clear and effectively communicates feedback, but here's how you can improve:

1. Expand on Details: Provide more specific examples of what made the materials unclear or what errors were present. Detail the impact of the workshop starting late.

2. Enhance Vocabulary: Use a wider range of vocabulary to express your points more precisely and variedly. For instance, instead of "informative," you might use "insightful" or "valuable."

3. Improve Grammar and Structure: Incorporate a variety of sentence structures to show grammatical range. For example, use complex sentences and passive constructions where appropriate.

4. Add More Interaction: Include suggestions for interactive elements with specific examples, like group discussions or case studies, to show a deeper understanding of engagement.

5. Polish Transitions: Use a variety of linking phrases to improve the flow between paragraphs and ideas.

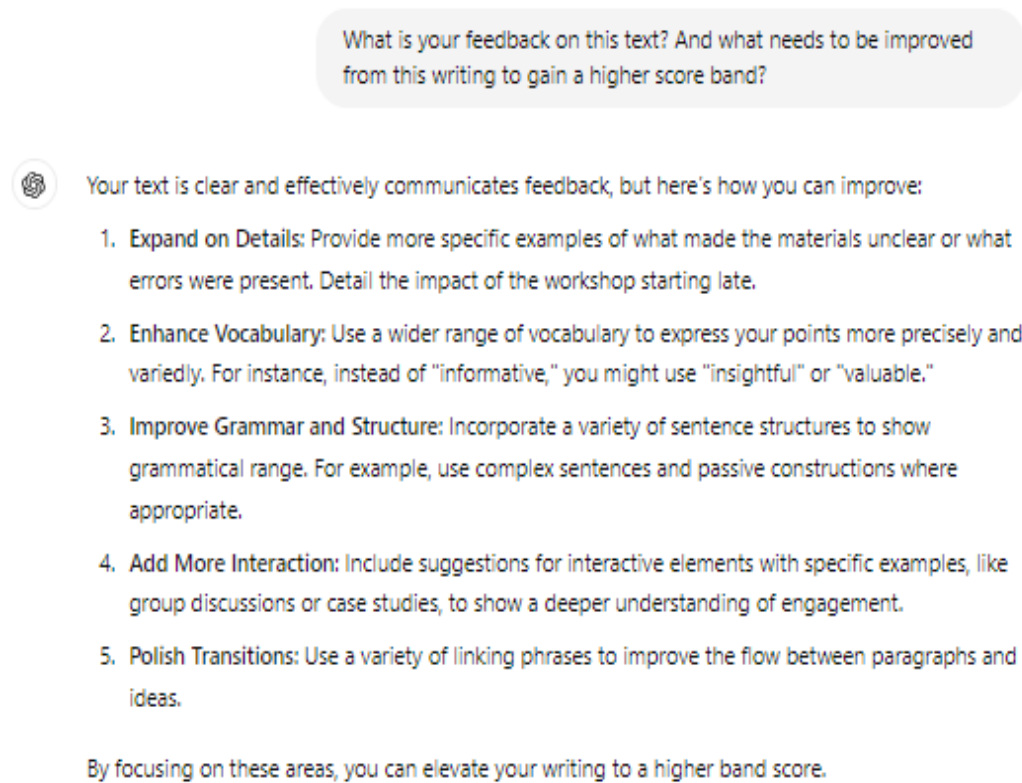By focusing on these areas, you can elevate your writing to a higher band score.

Figure 7. Short Feedback for the Mockup Evaluation

Most of the time, ChatGPT gave its feedback and things to improve for each scoring criterion, as in Figure 6. However, a few instances showed that ChatGPT did not go over each criterion, but rather gave overall feedback and things to improve, resulting in shorter responses. The feedback format in Figure 7 is actually more similar to the findings from Sutrisno (2023), that ChatGPT did not specify the feedback based on the scoring criteria but on the overall angle. However, the researcher is still unable to conclude the reason why this format did not become the main template. A possible influence is probably due to the request to provide the things that need to be improved to achieve a higher band.

In addition to feedback, the researcher also asked about the areas to improve. ChatGPT could provide the things to improve from the provided mockups. The content of the things to improve that it gave mainly about adding specific details (the context or background of the issue), using more linking verbs or transitional phrases, using a wider range of vocabulary, and using more complex sentences or a wider variety of sentence structure, as in Examples [16] and [17].

[16] Band 7 prompt 4: **Task Achievement** – Suggestions for Improvement: To enhance Task Achievement, consider providing more specific details if the prompt requires it…

Coherence and Cohesion – Suggestions for Improvement: …by using a wider range of linking phrases and ensuring smooth transitions between paragraphs…

Lexical Resource – Suggestions for Improvement: To score higher in Lexical Resource, use a wider range of vocabulary and avoid repetition.

For example, instead of using "request" multiple times, you could use synonyms like "inquiry" or "need."
Grammatical Range and Accuracy – Suggestions for Improvement: …focus on using a variety of sentence structures, including more complex sentences…

[17] Band 6 prompt 1: **Task Achievement** – Suggestions for Improvement: …Providing more specific examples might make your argument more compelling.
Coherence and Cohesion – Suggestions for Improvement: …Consider adding a linking phrase between the description of the issues…
Lexical Resource – Suggestions for Improvement: …you could use a wider range of vocabulary to avoid repetition. Words like "distraction," "interference," and "productivity" can be used to demonstrate a richer language ability.
Grammatical Range and Accuracy – Suggestions for Improvement: varying sentence structures a bit more… incorporating more complex sentences could enhance the overall quality.

As seen in Examples [16] and [17], ChatGPT could mention some aspects for the mockup improvements. Even though the presented mockups were from different score bands, ChatGPT gave similar points about what to improve. Nevertheless, ChatGPT seemed to add a bit more details about what to improve. For the task achievement, it requested more details. In the coherence and cohesion part, it suggested more linking phrases. In the lexical resource part, ChatGPT informed the use of wider vocabulary to avoid repetition. However, it also gave some suggestions of words people can use instead to improve the letter. Similar to the other two criteria, for the grammar part, it suggested varying sentence structure and using more complex sentences. Even so, it did not specify whether 'more complex sentences' meant using a higher number of complex sentences, whose structure is at least one independent and one dependent clause (Andriani & Bram, 2021), or increasing the complexity of the sentences. This ambiguity might confuse learners who are still undergoing the learning process to acquire English proficiency. ChatGPT's struggle to provide more in-depth feedback on how to improve the mockups is aligned with Sun (2023) that AI tools have limitations in providing thorough feedback.

In addition to giving word examples for better scores in the vocabulary criterion, ChatGPT sometimes gave revised examples with its suggestions taken into account. Among fifteen mockups, six were given the revised examples, as in Figure 8.

Here's a revised version of your letter incorporating these suggestions:

Dear [Landlord's Name],

I hope this letter finds you well. I am writing to provide feedback on the apartment I recently rented from you. Overall, I am pleased with the apartment's size and its convenient location, which is ideal for my daily commute.

However, there are a couple of issues that require your attention. Firstly, the heating system appears to be malfunctioning; it struggles to maintain a comfortable temperature and remains insufficient, especially during the colder evenings. Additionally, the kitchen sink has been leaking since my arrival, resulting in noticeable water damage underneath it.

Could you please arrange for a technician to inspect and repair both the heating system and the sink at your earliest convenience? These improvements would greatly enhance my comfort and overall satisfaction with the apartment.

Thank you for your prompt attention to these matters.

Figure 8. Revised Version of the Mockup

Learners can improve their writing when given feedback on their original writing's strengths and areas of improvement. However, sometimes just knowing the things to improve is still not enough to demonstrate better writing. In six responses, ChatGPT gave the revised version of the mockups that incorporated the suggestions it gave. This extra mile can help users visualize its suggestions for the betterment of their writing. This could further showcase the benefit of GenAI tools as a tutor, which not only gives feedback but also actually shows the revised writing. Though previous studies mentioned ChatGPT's ability to provide feedback and score bands, the revised version had not yet been explored in Sutrisno (2023). Though providing the revised version is unique, the ability to deliver the revised version itself is not rare for ChatGPT as a GenAI tool, as discussed by Wang et al. (2024) and Yan (2023). Therefore, if learners ever needed the revised letter to illustrate the initially given feedback, they could give the command to ChatGPT.

In summary, ChatGPT evaluates IELTS Writing Task 1 mockup answers by scoring each criterion individually before providing an overall band score. While its scoring aligns with the intended bands in 40% of cases, the majority of discrepancies involve either higher or lower estimations, though ChatGPT also provided the reasoning. ChatGPT's evaluations often highlight strengths and areas for improvement across scoring criteria, offering actionable suggestions such as incorporating more specific details, using varied vocabulary, and employing complex sentence structures. Additionally, its ability to present revised examples further enhances its value as a learning tool, illustrating feedback in practice. However, occasional inconsistencies in the depth of feedback and ambiguity in suggestions like "more complex sentences" could affect its effectiveness for learners. One key concern is the inconsistency in ChatGPT's scoring. This may happen because its evaluation depends on patterns from its training data, and it can change depending on the prompt and the text's main features. This lack of consistency can confuse learners who rely on GenAI tools to monitor their own progress, possibly misjudging their writing ability. As a result, these scoring

problems can negatively affect learners' efforts to improve independently, which is an important part of developing autonomous learning skills. These findings suggest that while ChatGPT is a capable evaluator and tutor, there is room for improvement to enhance its consistency, precision, and clarity in supporting IELTS preparation.

Conclusion

This study investigated how ChatGPT generated and evaluated IELTS Writing Task 1 answer mockups. For generating mockups, ChatGPT provided answers for score bands 6, 7, and 8 in a single response, varying in length, detail, and complexity according to the score band. It consistently followed task prompts while conforming to the writing criteria. However, some special cases appeared in low occurrences, such as not adhering to the word limit, repetitive ideas, occasional minor grammatical ambiguities, and limited differentiation in idea transitions. When evaluating mockups, ChatGPT assigned scores to each criterion and provided detailed feedback on strengths and areas for improvement, often supported by revised examples. While this feedback was mostly comprehensive, it sometimes lacked depth or clarity, particularly in suggestions like "more complex sentences". Another intriguing finding was the discrepancies between ChatGPT's evaluation scores and the intended bands of the generated mockups. That raised concerns about the consistency and accuracy of the scoring, whether the generated mockups or the evaluation got the most accurate bands. These findings suggest that while ChatGPT demonstrates potential as a learning aid, its outputs should be supplemented with human oversight to ensure accuracy and variability.

This study was still limited in some aspects. This study only used a single GenAI tool (ChatGPT), which rendered the findings ungeneralizable. It also did not include the human factor in the process. By including human-made mockups and human-made evaluations, the findings could be triangulated and generate richer insights on the topics. Additionally, the study focused solely on IELTS General Writing Task 1. Future research can expand this research by using other IELTS tasks or writing assignments and including multiple GenAI tools for comparison and incorporating the human aspect, whether in the form of learners or expert reviewers, to yield more robust insights. This study implies that GenAI tools like ChatGPT can provide a more accessible option for IELTS preparation by offering learners supplementary mockups and feedback, language learners and teachers need to critically evaluate the responses when utilizing GenAI in writing process, and further evaluation and development are necessary for the future development of GenAI tools to improve accuracy and consistency in writing assessments.

References

Alberth. (2023). The use of ChatGPT in academic writing: A blessing or a curse in disguise? *Teflin Journal*, *34*(2), 337–352. https://doi.org/10.15639/teflinjournal.v34i2/337-352

Alexopoulou, T., Michel, M., Murakami, A., & Meurers, D. (2017). Task effects on linguistic complexity and accuracy: A large-scale learner corpus analysis employing natural language processing techniques. *Language Learning*, *67*(1), 180–208. https://doi.org/10.1111/lang.12232

Andriani, E., & Bram, B. (2021). Syntactic analysis of sentence patterns and types of BBC news articles. *LiNGUA 16*(2), 25-34. https://doi.org/10.18860/ling.v16i1.10832

Andriani, E., & Ena, O. T. (2022). Teacher's use of code-switching types and functions in teaching TOEFL and IELTS preparation classes. *Pedagogy: Journal of English Language Teaching*, *10*(2), 153-168. https://doi.org/10.32332/joelt.v10i2.4379

Ariyanti, A., & Anam, S. (2021). Technology-enhanced paraphrasing tool to improve EFL students' writing achievement and enjoyment. *Journal of English Language Teaching and Linguistics*, *6*(3), 715-726. https://doi.org/10.21462/jeltl.v6i3.698

Assad, A. (2024). Exploring the use of ChatGPT among media students in Egypt: Opportunities and challenges. *Online Journal of Communication and Media Technologies*, *14*(2), 1-14. https://doi.org/10.30935/ojcmt/14416

Athanassopoulos, S., Manoli, P., Gouvi, M., Lavidas, K., & Komis, V. (2023). The use of ChatGPT as a learning tool to improve foreign language writing in a multilingual and multicultural classroom. *Advances in Mobile Learning Educational Research*, *3*(2), 818–824. https://doi.org/10.25082/amler.2023.02.009

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*, *57*, 1-6. https://doi.org/10.1016/j.asw.2023.100745

Budjalemba, A. S., & Listyani, L. (2020). Factors contributing to students difficulties in academic writing class: Students perceptions. *UC Journal: ELT, Linguistics and Literature Journal*, *1*(2), 135–149. https://doi.org/10.24071/uc.v1i2.2966

Buriak, J. M., Akinwande, D., Artzi, N., Brinker, C. J., Burrows, C., Chan, W. C. W., Chen, C., Chen, X., Chhowalla, M., Chi, L., Chueh, W., Crudden, C. M., Di Carlo, D., Glotzer, S. C., Hersam, M. C., Ho, D., Hu, T. Y., Huang, J., Javey, A., … Ye, J. (2023). Best practices for using AI when writing scientific manuscripts. *ACS Nano*, *17*(5), 4091–4093. https://doi.org/10.1021/acsnano.3c01544

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education.

Emran, A. Q., Mohammed, M. N., Saeed, H., Abu Keir, M. Y., Alani, Z. N., & Mohammed Ibrahim, F. (2024). Paraphrasing ChatGPT answers as a tool to enhance university students' academic writing skills. *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems, ICETSIS 2024*, 501–505. https://doi.org/10.1109/ICETSIS61505.2024.10459386

Etikan, I. & Bala, K. (2017). Sampling and Sampling Methods. *Biometrics & Biostatistics International Journal*, *5*(6), 215-217. https://doi.org/10.15406/bbij.2017.05.00149

Fitria, T. N. (2021). QuillBot as an online tool: Students' alternative in paraphrasing and rewriting of English writing. *Englisia: Journal of Language, Education, and Humanities*, *9*(1), 183-196. https://doi.org/10.22373/ej.v9i1.10233

Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing assistant's impact on English language learners. *Computers*

and Education: Artificial Intelligence, *3*, 1-7. https://doi.org/10.1016/j.caeai.2022.100055

Ginting, P., Muda Batubara, H., & Hasnah, Y. (2023). Artificial intelligence powered writing tools as adaptable aids for academic writing: Insight from EFL college learners in writing final project. *International Journal of Multidisciplinary Research and Analysis,* 6(10), 4640-4650. https://doi.org/10.47191/ijmra/v6-i10-15

Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation,* 5(1), 37-45. https://doi.org/10.61414/jeti.v5i1.103

Huang, A. Y. Q., Lu, O. H. T., & Yang, S. J. H. (2023). Effects of artificial intelligence–enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education*, *194*, 104684. https://doi.org/10.1016/j.compedu.2022.104684

IELTS. (https://ielts.org/) (accessed 2021-08-20).

Imran, M., & Almusharraf, N. (2023). Analyzing the role of ChatGPT as a writing assistant at higher education level: A systematic review of the literature. *Contemporary Educational Technology,* *15*(4), 1-14. https://doi.org/10.30935/cedtech/13605

Jeanjaroonsri, R. (2023). Thai EFL learners' use and perceptions of mobile technologies for writing. *LEARN Journal: Language Education and Acquisition Research Network,* *16*(1), 169-193. https://so04.tci-thaijo.org/index.php/LEARN/index

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal,* *54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Koraishi, O. (2023). Teaching English in the age of AI: Embracing ChatGPT to optimize EFL materials and assessment. *LET Journal: Language Education & Tehcnology, 3*(1), 55-72. http://langedutech.com

Kurniasari, M. D. (2017). Tolerating structural ambiguity in grammar learning. *IJIET (International Journal of Indonesian Education and Teaching)*, *1*(1), 85–94. https://doi.org/10.24071/ijiet.v1i1.330

Kurniati, E. Y., & Fithriani, R. (2022). Post-graduate students' perceptions of Quillbot utilization in English academic writing class. *Journal of English Language Teaching and Linguistics*, *7*(3), 437-451. https://doi.org/10.21462/jeltl.v7i3.852

Lavidas, K., Voulgari, I., Papadakis, S., Athanassopoulos, S., Anastasiou, A., Filippidi, A., Komis, V., & Karacapilidis, N. (2024). Determinants of humanities and social sciences students' intentions to use artificial intelligence applications for academic purposes. *Information, 15*(6), 314. https://doi.org/10.3390/info15060314

Le, C., & Mohd, T. K. (2024). A comparative study of text comprehension in IELTS reading exam using GPT-3. *International Journal of Computer Science and Information Technology*, *16*(1), 45–52. https://doi.org/10.5121/ijcsit.2024.16104

Lichtinger, E. (2018). Gap between self-efficacy and college students' writing skills. *Journal of College Reading and Learning*, *48*(2), 124–137. https://doi.org/10.1080/10790195.2017.1411213

Lin, M. P.-C., & Chang, D. (2020). Enhancing post-secondary writers' writing skills with a chatbot: A mixed-method classroom study. *Journal of Educational Technology & Society*, *23*(1), 78–92. https://www.jstor.org/stable/26915408

Lin, X. (2023). Exploring the role of ChatGPT as a facilitator for motivating self-directed learning among adult learners. *Adult Learning*, *35*(3), 156–166. https://doi.org/10.1177/10451595231184928

Maghamil, M. C., & Sieras, S. G. (2024). Impact of ChatGPT on the academic writing quality of senior high school students. *Journal of English Language Teaching and Applied Linguistics, 6*(2), 115-128. https://doi.org/10.32996/jeltal

Maharani, A. A. P., & Santosa, M. H. (2021). The implementation of process approach combined with Screencast-o-matic and Bookcreator to improve students' argumentative writing. *LLT Journal: Journal on Language and Language Teaching*, *24*(1), 12–22. https://doi.org/10.24071/llt.v24i1.2516

Marzuki, Widiati, U., Rusdin, D., Darwin, & Indrawati, I. (2023). The impact of AI writing tools on the content and organization of students' writing: EFL teachers' perspective. *Cogent Education*, *10*(2), 1-17. https://doi.org/10.1080/2331186X.2023.2236469

Meniado, J. (2023). The impact of ChatGPT on English language teaching, learning, and assessment: A rapid review of literature. *Arab World English Journal*, *14*(4), 3–18. https://doi.org/10.24093/awej/vol14no4.1

Mogavi, R., Deng, C., Juho Kim, J., Zhou, P., D. Kwon, Y., Hosny Saleh Metwally, A., Tlili, A., Bassanelli, S., Bucchiarone, A., Gujar, S., Nacke, L. E., & Hui, P. (2024). ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters' utilization and perceptions. *Computers in Human Behavior: Artificial Humans*, *2*(1), 1-20. https://doi.org/10.1016/j.chbah.2023.100027

OpenAI. (https://openai.com/) (2024-08-20).

Rahman, B. I. R, Laiya, R. E., & Sarumaha, M. S. (2023). Navigating English writing proficiency tests in the era of artificial intelligence. *Journal of English Education and Teaching, 7*(3), 480-498. https://doi.org/10.33369/jeet.7.3.480-498

Shaib, C., Elazar, Y., Li, J. J., & Wallace, B. C. (2024). Detection and measurement of syntactic templates in generated text. *Computation and Language,* 1-16. https://doi.org/10.48550/arXiv.2407.00211

Silva, C. (2017). Research design - The new perspective of research methodology. *British Journal of Education, Society & Behavioural Science*, *19*(2), 1–12. https://doi.org/10.9734/bjesbs/2017/30274

Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, *14,* 1-14. https://doi.org/10.3389/fpsyg.2023.1260843

Su, Y., Lin, Y., & Lai, C. (2023). Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*, *57*, 100752. https://doi.org/10.1016/j.asw.2023.100752

Sun, T. (2023). The potential use of generative AI in ESL writing assessment: A case study of IELTS writing tasks. *Irish Journal of Technology Enhanced Learning, 7*(2), 42-51. http://www.ilta.ie/

Susanto, A. I. (2023). ChatGPT vs. human IELTS tutor: A comparative study of IELTS task 2 writing sample. *20th Technology for Second Language Learning Conference.* https://apling.engl.iastate.edu/conferences/technology-for-second-language-learning-conference/tsll-2023/abstracts/

Sutrisno, D. B. (2023). Practicing IELTS writing for L2 writers with ChatGPT; An exploratory self-study. *English Department of UMMU Journal (EDU Journal*, *3*(2), 41-55. http://www.jurnal.ummu.ac.id/index.php/edu

Thorp, H. (2023). ChatGPT is fun, but not an author. *Science*, *379*, 313. https://doi.org/10.1126/science.adg7879

Triana, H., Abustang, P. B., Utomo, E., Rakhman, G. G. F., & Fahrurrozi, F. (2023). Assessment evaluation using solo taxonomy for measuring levels of critical thinking skills: PYP international baccalaureate case study. *Prisma Sains: Jurnal Pengkajian Ilmu Dan Pembelajaran Matematika Dan IPA IKIP Mataram*, *11*(2), 581-595. https://doi.org/10.33394/j-ps.v11i2.7894

Utami, S. P. T., Andayani, Winarni, R., & Sumarwati. (2023). Utilization of artificial intelligence technology in an academic writing class: How do Indonesian students perceive? *Contemporary Educational Technology*, *15*(4), 1-13. https://doi.org/10.30935/cedtech/13419

Van Horn, K. R. (2024). ChatGPT in English language learning: Exploring perceptions and promoting autonomy in a university EFL context. *TESL-EJ*, *28*(1), 1-26. https://doi.org/10.55593/ej.28109a8

Vicente-Yagüe-Jara, M. I., López-Martínez, O., Navarro-Navarro, V., & Cuéllar-Santiago, F. (2023). Writing, creativity, and artificial intelligence. ChatGPT in the university context. *Comunicar*, *31*(77), 47–57. https://doi.org/10.3916/C77-2023-04

Wafa, N., Samia, M., Mohamed, H., Amine, E., & Fatiha, K. B. (2024). Innovating narratives or stifling creativity? Assessing the impact of generative pre-trained transformer (GPT) on science fiction writing skills among master students in Algeria. *Educational Administration: Theory and Practice*, *30*(6), 1494–1507. https://doi.org/10.53555/kuey.v30i6.5526

Wale, B. D., & Kassahun, Y. F. (2024). The transformative power of AI writing technologies: Enhancing EFL writing instruction through the integrative use of Writerly and Google Docs. *Human Behavior and Emerging Technologies*, *2024,* 1-15. https://doi.org/10.1155/2024/9221377

Wang, C., Li, Z., & Bonk, C. (2024). Understanding self-directed learning in AI-Assisted writing: A mixed methods study of postsecondary learners. *Computers and Education: Artificial Intelligence*, *6,* 1-10. https://doi.org/10.1016/j.caeai.2024.100247

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education*, *19*(7), 1-18. https://doi.org/10.29333/ejmste/13272

Wilson, J., Myers, M., & Potter, A. (2022). Investigating the promise of automated writing evaluation for supporting formative writing assessment at scale. *Assessment in Education: Principles, Policy & Practice*, *29*, 1–17. https://doi.org/10.1080/0969594X.2022.2025762

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/10.1177/0735633119830764

Xiao, Y., & Zhi, Y. (2023). An exploratory study of EFL learners' use of ChatGPT for language learning tasks: Experience and perceptions. *Languages*, *8*(3), 1-12. https://doi.org/10.3390/languages8030212

Yan, D. (2023). How ChatGPT's automatic text generation impact on learners in a L2 writing practicum: An exploratory investigation. *Education and Information Technologies, 28*, 13943–13967. https://doi.org/10.1007/s10639-023-11742-4

Zhang, Z. (2024). ChatGPT: Enhancing learner agency and multilingual practices in English writing class. *Journal of Education, Humanities and Social Sciences 26*, 1153-1160. https://doi.org/10.54097/2cwdzn44

Zhao, X., Hu, J., & Cox, A. (2024). Incorporating artificial intelligence into student academic writing in higher education: The use of Wordtune by Chinese international students. *Proceedings of the 57th Hawaii International Conference on System Sciences,* 2726-2735. https://hdl.handle.net/10125/106712

Zhou, D., & Zhang, Y. (2024). Red AI? Inconsistent responses from GPT models on political issues in the US and China. *Computation and Language,* 1-21. https://doi.org/10.48550/arXiv.2312.09917