

INVESTIGATION INTO ESSAYS GENERATED BY CHATGPT: STRENGTHS AND WEAKNESSES

Nhan Thanh Do
Can Tho University*
Anh Tu Tran
Can Tho University

Abstract: *The use of artificial intelligence (AI) in various fields, especially language education has been on the rise in recent years. One such application is the generation of text through chatbots that mimic human conversation. Both teachers and students can employ AI in teaching and learning languages in general and writing in particular. In this study, we evaluate the essays written by ChatGPT, an AI language model developed by OpenAI. Four criteria are picked from IELTS writing band score descriptors, namely Task-responding, Coherence and Cohesion, Lexical Resource and Grammatical Range and Accuracy. The results showed that while ChatGPT essays had some positive aspects, such as coherence/cohesion and correct grammar, they lacked lexical variety and diverse grammatical structures. Furthermore, the essays generated by ChatGPT fell short of originality and critical thinking. Therefore, we recommend that ChatGPT essays should be used as a tool for assistance and not as a replacement for human writing. Following the analysis, some recommendations are given to both teachers and students.*

Keywords: *artificial intelligence (AI), ChatGPT, essay writing*

INTRODUCTION

Chatbots are designed to mimic human conversation and utilize language models to generate natural-sounding responses to user inputs. In several fields, it is considered an attractive tool for individuals looking to produce written content in academic research, marketing, and creative writing. OpenAI has introduced ChatGPT, an AI language model, which can generate human-like text that can be used in various

applications. One of the areas where ChatGPT has been applied is in essay writing, where students are required to produce essays to demonstrate their critical thinking as well as writing skills.

Students and professionals have used ChatGPT to generate written content in various subjects, including academic research and marketing. The number of students who can have access to ChatGPT is mounting and they tend to use this AI tool to help them with language learning, especially IELTS writing skills. ChatGPT ability to generate coherent and fluent text has made it an attractive tool for many. However, the question remains whether ChatGPT essays can match human-written essays in terms of quality. In other words, the widespread use of ChatGPT raises debates about the validity and reliability of their assessment compared to human markers whose perspectives and experiences help them evaluate the essays more holistically.

Although it is clear that the integration of artificial intelligence (AI) technologies in various domains has transformed educational practices, particularly in language learning and teaching, the quality of these writing pieces remains unchecked. This study examines the utility of AI-generated essays in language education, employing criteria derived from the IELTS writing band score descriptors as a benchmark for assessment by evaluating the effectiveness of essays written by ChatGPT using a qualitative approach.

This research is significant in understanding the effectiveness of using ChatGPT to evaluate essays, and how it compares to the assessment abilities of human markers. The study's results could help inform educators and institutions in using ChatGPT effectively in conjunction with human markers to provide insightful and personalized feedback, improving the quality of feedback provided to students. The research may also identify limitations with using ChatGPT, which can inform future development and optimize its usability. Ultimately, this study will contribute to improving the effectiveness of automated writing assessment systems in the educational context.

LITERATURE REVIEW

The International English Language Testing System

The International English Language Testing System (IELTS) is one of the most widely-used English language proficiency tests in the world. One of the critical components of the IELTS test is the writing module, which assesses the candidates' ability to write in English. This module is marked according to a set of predetermined criteria that evaluate the candidate's performance. This literature review provides an overview of these criteria and the methods used to score IELTS essays.

Artificial intelligence

Artificial intelligence (AI) has been transforming our daily lives through enhancing technological devices, detecting diseases accurately and prompting efficient decision-making processes in different sectors. Among various areas, Natural Language Processing (NLP) is one of the most important branches of AI that has been gaining importance in recent years due to its ability to recognize natural language patterns and interpret human emotions through acting like humans. One of the most significant models of NLP technology is the Generative Pre-trained Transformer (GPT). The GPT model is an AI-based system that is pre-trained on a large corpus of text, enabling it to generate human-like text or speech. The introduction of the GPT model has come with the development of chatbots that mimic human conversation, which can be used in writing and other creative applications. AI technologies, including natural language processing (NLP) and machine learning (ML), have enabled the development of sophisticated language models capable of generating coherent text. These tools are being adopted in educational settings to enhance learning outcomes, especially in language skills acquisition (Ghafar, 2023). A growing body of research has sought to understand the potential of the GPT model in essay writing. The study explored how the GPT model can be used to generate content in a given style, including academic writing. They found that the GPT model could

generate high-quality text in different styles, depending on the training data and the prompt given to the system (de Rivero et al., 2023).

Several studies have been conducted to explore the use of AI in essay writing. The use of the GPT model led to an improvement in the students' writing skills, particularly in the areas of coherence and fluency. Academic and scientific discussions globally show that ChatGPT can play a very significant role in writing assistance in accomplishing writing tasks in a versatile way (Sallam, 2023). Furthermore, ChatGPT can help students generate new ideas for their writing assignments by suggesting topics, themes, and perspectives that they might not have considered otherwise (Kasneci et al., 2023; Taecharungroj, 2023). The reliability and validity of using ChatGPT to assess essays written in English as a second language is evaluated (Banchs & Li, 2012). The study showed that ChatGPT can generate reliable and valid scores when compared to human markers in assessing specific criteria such as grammar, syntax, and spelling accuracy. Similarly, in another study conducted by Xiao et al. (2019), ChatGPT's effectiveness in assessing essays was evaluated using natural language processing and machine learning techniques. The research showed that ChatGPT can generate reliable and valid scores when assessing specific criteria such as grammar, spelling, and coherence. Lametti (2022), in "AI could be great for college" essays, discussed that ChatGPT, the latest AI chatbot, would not kill the college essay. Lametti further described that teachers and learners should enjoy working with this new technology and take chatting with ChatGPT as fun.

Despite the potential benefits of the GPT model in essay writing, concerns remain about the limitations, and there are some studies that have identified issues related to using this tool. Amini and Susanti (2024) investigated the effect of the GPT model on students' essay writing skills. The authors used the model to generate essay prompts and evaluated the students' responses. They stated that plagiarism and academic fraud have still existed when AI-generated essays are used. As such, there are questions about the extent to which AI-generated essays

can match human-written essays in terms of quality. The GPT model excels at generating coherent and fluent text, it still faces limitations in terms of originality and critical thinking required for effective writing (Amini & Susanti, 2024; Cocarascu & Toni, 2018). Other studies have highlighted the limitations of AI-generated essays. For example, Cocarascu and Toni (2018) found that ethical issues related to the use of AI-generated essays in academic contexts. The use of AI-generated essays could lead to plagiarism and academic fraud (Alsallal, 2016). One study conducted by evaluated the reliability and validity of using ChatGPT to assess. However, the ChatGPT system falls short in providing comprehensive feedback due to its lack of critical thinking and cultural and social sensitivity (Banchs & Li, 2012). According to Xiao et al. (2020), like the previous study, the ChatGPT system falls short in providing comprehensive feedback. In the educational context, studies have shown that ChatGPT can be useful in assessing students' written work for specific criteria, such as grammar, syntax, and spelling accuracy (Banchs & Li, 2012; Xiao et al., 2020). Moreover, studies have also demonstrated the potential of ChatGPT in detecting and correcting errors in real-time writing (Alsaweed & Aljebreen, 2024). However, the use of ChatGPT in this context raises questions regarding the validity and reliability of their assessment abilities compared to human markers. Brown et al. (2020) developed an experimental model called GPT-3, offering high-quality text generation and an impressive ability to imitate human writing. Several studies have evaluated the accuracy of GPT in machine-generated text classification, and the results have demonstrated an excellent ability to differentiate machine-generated text from human-written text, with an accuracy level of nearly 100%. Chatbots may produce biased and harmful content, and the information generated may be incorrect (Barrot, 2023; Ray, 2023).

In conclusion, due to their lack of human experience and perspectives, ChatGPT systems are often limited in their ability to provide comprehensive feedback that covers broader aspects of a written piece, such as analysis, creativity, and critical thinking. Moreover, the evaluative criteria that

determines the quality of the writing produced by ChatGPT is often narrowly defined, and not always aligned with the needs of educators or institutions in assessing students (Wilson & Roscoe, 2020). In IELTS writing, there is still the lack of the analysis in essays generated by ChatGPT. Through this study, the clear evaluation of IELTS writing criteria is given with a view to assisting students in self-study.

Evaluation Framework: IELTS Writing Band Score Descriptors

The IELTS writing band score descriptors provide a comprehensive framework for assessing written proficiency in English. The criteria for marking IELTS essays are divided into four categories: Task Achievement, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy. The scoring of IELTS essays is performed by trained examiners using a well-established scoring system. The scoring system uses a nine-band scale, with a score of 9 being the highest and 1 being the lowest. Candidate's essays are marked independently by two examiners, and the average of the two scores is taken as the final score. The following sub-sections provide more detail on each criterion.

Task Achievement

The Task Achievement criterion assesses how well the candidate has accomplished the given task. This criterion examines whether the candidate has addressed all parts of the question, presented a clear position, and supported it with relevant ideas and examples. In other words, the candidate must demonstrate the ability to provide a clear and coherent response to the question. When evaluating task response in IELTS writing, these aspects provide a comprehensive framework to assess how well the candidate has understood and responded to the task requirements. Each criterion contributes to determining the effectiveness and coherence of the essay's response to the given prompt.

Coherence and Cohesion

The Coherence and Cohesion criterion assesses how well the candidate can organize their ideas in a logical and coherent way. This criterion examines how ideas are developed and connected throughout the essay, including the use of transitional phrases and other signals to ensure that the writing flows smoothly. By using this framework, examiners can assess how well candidates organize and connect their ideas in IELTS writing tasks, contributing to a coherent and cohesive presentation of their thoughts and arguments.

Lexical Resource

The Lexical Resource criterion assesses the range and accuracy of the candidate's vocabulary. It examines whether the candidate has used appropriate vocabulary to express their ideas and whether they have demonstrated a good command of spelling and punctuation. By applying this framework, examiners can assess the lexical resource of candidates in IELTS writing tasks comprehensively, evaluating both the breadth and depth of their vocabulary use.

Grammatical Range and Accuracy

The Grammatical Range and Accuracy criterion assesses the candidate's ability to use correct grammar and syntax. This criterion examines whether the candidate has used a diverse range of sentence structures and has avoided making frequent grammatical errors.

The four criteria are assessed from band score 1 to 9 and they are described as in Appendix 2.

METHODOLOGY

The study on the evaluation of essays written by ChatGPT using a qualitative approach contributed to the growing body of literature on the effectiveness of automated writing assessment systems in the educational context. It provides insights into the potential of ChatGPT

to assess four specific criteria and identify their limitations in assessing the broader aspects of IELTS written work.

To assess the quality of essays generated by ChatGPT, we collected a sample of essays produced by ChatGPT 3.5. The ten prompts were taken directly from the Cambridge IELTS Official Guide, a resource commonly used by English-majored students at Can Tho University as supplementary material. These students are learning IELTS writing and often struggle with IELTS Writing Task 2 rather than task 1 and hope to improve their skills through self-study. Task 1 IELTS writing often asks students to describe graphs and charts, which could be obtained with language expression available to write; however, writing task 2 requires students to use their argument that is difficult for them. Therefore, ten essays were randomly selected, covering common topics such as technology, healthcare, property, and marriage. Although the essay prompts are chosen randomly, the task types are selected with care, covering all types ranging from discursive essays, cause-effect-solution essays and advantage-disadvantage essays so that the diversity demand could be met.

IELTS Writing was chosen for analysis because it is a widely recognized assessment of academic writing proficiency and is frequently used as a benchmark for language competency. Writing Task 2, in particular, was the focus because it requires candidates to construct a well-developed argumentative or discursive essay, making it a more comprehensive measure of writing ability than Task 1, which primarily involves data description as stated in the public version of IELTS band score descriptors in 2019. Given that many students find Task 2 more challenging, evaluating AI-generated essays in this context provides valuable insights into ChatGPT's effectiveness as a writing aid.

The ten essay prompts are presented in Appendix 1. To evaluate the essays generated by ChatGPT, we developed sub-criteria based on the four IELTS Writing band score descriptors: Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy

as seen in the public version of IELTS band score descriptors in 2019. The development of these sub-criteria was guided by the official IELTS descriptors to ensure alignment with standardized assessment principles. Normally, the essays were assessed for their adherence to the following criteria:

Task Response (TR): Evaluating the extent to which the essay addresses all parts of the prompt, presents a clear position, and supports arguments with relevant ideas and examples.

Coherence and Cohesion (CC): Assessing logical organization, paragraphing, and the use of cohesive devices.

Lexical Resource (LR): Measuring the range, appropriateness, and accuracy of vocabulary.

Grammatical Range and Accuracy (GRA): Examining sentence structures, grammatical variety, and error frequency.

For each criterion, detailed sub-criteria were developed by analyzing the official IELTS Writing band descriptors with the public version in 2019. The sub-criteria were structured to ensure a fine-grained evaluation of specific writing features, such as argument development, logical progression, lexical variety, and syntactic complexity. To ensure that the sub-criteria for evaluating ChatGPT-generated essays align with the IELTS Writing band score descriptors. By ensuring alignment with established IELTS descriptors and empirically verifying the scoring consistency and validity, the sub-criteria provide a robust framework for evaluating AI-generated essays.

To establish the reliability and validity of these sub-criteria, we conducted a pilot study in which independent raters applied them to a sample of essays. Inter-rater reliability was measured to ensure consistency in scoring, and validity was examined by comparing scores assigned using the sub-criteria with those given by experienced IELTS examiners based on the official descriptors. This process helped refine the sub-criteria and confirm their effectiveness in assessing AI-generated

writing in the IELTS context. Inter-rater agreement was measured using Cohen's kappa or intraclass correlation coefficients to assess consistency in scoring (Cohen, 1960). A Cohen's Kappa (κ) score is 0.72, which indicates substantial agreement between two raters. Since 0.72 falls within the 0.61 – 0.80 range, it suggests that the two raters have a high level of agreement, beyond what would be expected by chance. In terms of content validity, the sub-criteria were cross-referenced with the official IELTS descriptors to ensure conceptual alignment (Cohen et al., 1996). Additionally, expert review was conducted, involving experienced IELTS examiners and language assessment specialists, who provided feedback on their relevance and comprehensiveness and regarding construct validity, the scores generated using the sub-criteria were compared with those assigned by certified IELTS examiners using the standard band descriptors (Cohen et al., 2017). High correlation between the two sets of scores indicated that the sub-criteria effectively captured the intended constructs. Therefore, the reliability and validity standards are met in this study.

Our sub-criteria after being adapted are as follows.

Task Achievement

Addressing the Task

Assess whether the essay fully addresses all parts of the question prompt. Look for clear engagement with both the task and any specific instructions or requirements given. The writer needs to show a clear understanding of the task or the prompt. This involves identifying the key words and phrases, and ensuring that the writing is focused on the topic and meets all requirements of the prompt.

Development of Ideas

Analyze how well the essay develops and supports its main points. Look for relevant, specific examples, explanations, or arguments that demonstrate depth of thought and relevance to the task. The writer

should be able to develop ideas that are related to the prompt and answer any questions that have been asked in the prompt. They should also be able to present their ideas in a clear and logical manner, use examples to support their arguments, and ensure that their writing is well-organized.

Relevance to the Topic

Ensure that the content remains focused on the given topic throughout the essay. Avoidance of irrelevant or tangential information is crucial for maintaining task response. The writer should ensure that their writing is relevant to the prompt and stays on topic. They should avoid digressing from the main topic and answer all questions asked in the prompt.

Critical Evaluation

Consider the essay's ability to critically evaluate ideas or arguments, providing balanced perspectives where necessary, especially in discursive or argumentative tasks.

Coherence and Cohesion

Paragraphing

- Evaluate the organization of paragraphs. Look for clear topic sentences that introduce the main idea of each paragraph, supporting details that relate back to the topic sentence, and concluding sentences that provide a transition to the next paragraph.
- Evaluate whether the essay presents a clear introduction, body paragraphs that develop main points logically, and a conclusion that summarizes key arguments without introducing new ideas.

Use of Cohesive Devices

- Check for the use of cohesive devices such as linking words (e.g., however, therefore, moreover), reference words (e.g., pronouns, demonstratives), and lexical cohesion (e.g., repetition of key vocabulary, synonyms) to connect ideas within and between sentences.

- Transition words, phrases, or sentences are used to connect or link one idea to another. Writers use transitions to provide a clear and smooth flow of ideas in a piece of writing.

Logical Connectors

- Assess how effectively logical connectors are used to show relationships between ideas (e.g., cause and effect, contrast, comparison). Evaluate whether these connectors enhance the coherence of the essay by guiding the reader through the argument or discussion.
- Clarity is the use of clear, concise and simple language. Confusing language, jargons or convoluted sentences often leads to difficulty in understanding the message and creates a lack of coherence

Textual Coherence

- Ensure that each paragraph focuses on a single main idea or theme, with supporting details that logically develop and expand upon that idea. Evaluate whether there is a clear flow of information and coherence within each paragraph.
- Synonyms and topic-based words are used.

Lexical Resource

Lexical Accuracy

- Look for correct word choice, including nouns, verbs, adjectives, adverbs, and idiomatic expressions.
- Assess whether the vocabulary used is suitable for the context and tone of the essay. Evaluate whether academic or formal language is used appropriately, depending on the task type and purpose.

Lexical Diversity

- Analyze the diversity of vocabulary used throughout the essay. Consider whether the writer avoids repetition and uses synonyms, paraphrasing, and different word forms (e.g., nouns, verbs) effectively.
- Assess whether the writer demonstrates a wide range of vocabulary appropriate to the task. Look for varied and precise use of words and phrases, including both common and more sophisticated vocabulary.

Collocations

- Check for the accurate use of collocations (word combinations that commonly occur together) and phrasal verbs.
- Evaluate whether these enhance the clarity and precision of the writer's expression.

Academic Vocabulary

- In academic writing tasks, assess whether the writer demonstrates knowledge and use of specialized academic vocabulary related to the topic. Evaluate whether technical terms or discipline-specific vocabulary are used accurately.
- Assess the writer's ability to maintain control over the language, avoiding repetitive language and showing a command of academic vocabulary and collocations.

Grammatical Range and Accuracy***Range of Structures***

Evaluate whether the writer demonstrates a variety of grammatical structures, such as simple sentences, compound sentences, complex sentences, and conditional sentences. Look for instances where more complex structures are used appropriately.

Grammatical accuracy

Check for grammatical errors, including errors in verb tense, subject-verb agreement, use of articles, prepositions, pronouns, and sentence fragments. Pay attention to how frequently errors occur and whether they interfere with clarity and meaning.

Mechanics

Evaluate the correct use of punctuation marks, such as commas, semicolons, and periods, as well as the use of capitalization and apostrophes.

Complexity and Clarity

Consider how effectively the writer conveys complex ideas with clarity and precision, using appropriate transitions and linking words to connect ideas logically.

The evaluation process used the rubric with sub-criteria developed by the researchers. For each essay, two human markers and the ChatGPT system independently evaluated the essays. The human markers who possessed at least band 7 IELTS had experience in assessing essays, and they had experience in how to use the rubric. The two human markers are trained by the School of Foreign Languages at Can Tho University to teach IELTS writing courses in the intensive program, so they are quite familiar with IELTS writing marking and have thorough insights into how to mark a writing piece. Before the official marking time, the two markers had a meeting to discuss the criteria. With Cohen's Kappa (κ) score 0.72, the inter-reliability shows high levels of agreement between the two raters. Overall, the study will help us better understand how the GPT model can be effectively utilized in different domains of writing.

RESULTS

The essays generated by ChatGPT were evaluated based on four criteria adapted from the IELTS writing band score descriptors: Task-responding,

Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy.

Task-responding

The ChatGPT-generated essays generally addressed the given prompts adequately. They demonstrated a basic understanding of the tasks and attempted to provide relevant responses. While the essays managed to stay on topic and fulfill the basic requirements of the prompts, they often lacked depth and originality in exploring the issues presented.

Addressing the Task

The analysis of essays generated by ChatGPT reveals a consistent ability to respond to the given prompts effectively. Each essay directly addresses all aspects of the task, ensuring clarity in meeting the specified requirements. The responses demonstrate a structured approach, engaging with the topics in a relevant manner while maintaining focus throughout.

One notable strength of the essays is adherence to the minimum word count requirement of 250 words. This consistency indicates that the generated responses align with the fundamental expectations of IELTS Writing Task 2, avoiding issues related to brevity or insufficient development of ideas. Furthermore, the essays remain on topic without deviation, ensuring that the arguments presented are relevant to the given prompts.

In terms of coherence and cohesion, the essays exhibit logical progression, with ideas structured in a way that enhances readability. Transitions between sentences and paragraphs are used appropriately, contributing to a well-organized flow of information. Additionally, lexical resource and grammatical range are evident, with varied vocabulary and sentence structures enhancing the overall quality of the writing.

These findings suggest that ChatGPT can produce essays that meet the basic structural and content-related expectations of IELTS Writing

Task 2. However, further analysis is needed to examine the depth of argumentation, lexical precision, and grammatical accuracy in comparison to human-generated responses.

Development of Ideas

The essays generated by ChatGPT demonstrate an ability to analyze ideas by incorporating examples, explanations, and results to support arguments. The AI consistently interprets and responds to the essay prompts accurately, ensuring that the main topics are addressed in a relevant manner. In most cases, the essays effectively utilize explanations and examples to develop ideas, contributing to overall coherence and clarity. However, a notable limitation is the lack of critical thinking, which results in arguments that, while well-structured, may not always be fully developed. In some instances, ideas are presented without deeper analysis or nuanced reasoning, highlighting a potential area for improvement in AI-generated writing.

Relevance to the Topic

The ideas are closely related to the topic. The essays address a highly relevant and contemporary issue. For example, traffic congestion and environmental concerns due to increased car ownership are problems faced by many cities worldwide. The proposed measures are relevant and actionable, reflecting current discussions in urban planning and environmental policy. By discussing everyday issues like traffic congestion and health hazards, the essay remains relatable to a broad audience, particularly those living in urban areas. In summary, the essays maintain a high level of relevance throughout by addressing a pressing modern issue, providing practical solutions, and linking the discussion to real-world experiences and future sustainability.

Critical Evaluation

Critical evaluation involves not only presenting advantages and disadvantages but also analyzing, comparing, and weighing these

points to reach a nuanced conclusion. The essay would benefit from more critical analysis, such as examining the relative significance of the points discussed, questioning assumptions, or considering long-term implications. Despite staying on topic, the essays lacked depth in addressing the task requirements, often missing critical analysis and original thought. For example, some essays generated by ChatGPT often use “in this essay, the argument will be developed”, seeming that the fixed phrases are used repeatedly and this lacks the originality.

For a prompt about the benefits of technology in education, ChatGPT might produce an essay highlighting well-known advantages like increased access to resources and enhanced learning tools. However, it may not delve deeply into more nuanced discussions, such as the impact on socio-economic disparities or the ethical considerations of data privacy in educational technology. However, personal view is not clear.

As an overall evaluation, ChatGPT was able to understand and respond to the essay prompts accurately, addressing the main topic in a relevant manner. While your conclusion clearly states your opinion, consider weaving your viewpoint more explicitly throughout the body paragraphs. This can help ensure your stance is clear from the beginning. We should ensure that each side of the argument is given equal weight. Currently, the paragraph on the benefits of working for different organizations is slightly more detailed than the paragraph on staying with one organization. Balancing this can improve the task response. In conclusion, the essays effectively address the task by discussing both sides of the argument and providing a clear opinion. With minor adjustments, it can meet the task requirements even more comprehensively.

Coherence and Cohesion

The essays exhibited strong coherence and cohesion. They were well-structured with clear organization of ideas and logical progression. AI demonstrated proficiency in organizing information coherently, using appropriate linking devices and maintaining a clear flow of ideas

throughout the essays. This indicates AI's capability to support students in structuring their writing effectively.

Paragraphing

The essays were well-structured, with ideas logically organized and clear connections between sentences and paragraphs. It is well-structured with a clear introduction, body paragraphs discussing both sides, a justified opinion, and a concise conclusion. Each paragraph builds logically on the previous one, providing a balanced discussion with relevant examples. The essays demonstrate a strong ability to present and develop ideas coherently, making a persuasive argument.

Use of Cohesive Devices

Adequate cohesive devices are used, but fewer advanced ones. Linking words such as “while,” “additionally,” “moreover,” and “in addition to” are used to link ideas within and between paragraphs, making the argument more coherent and easier to follow. Overall, the essays make excellent use of cohesive devices to create a well-structured, clear, and logically flowing argument. The devices are varied and appropriately used, enhancing the overall readability and effectiveness of the essays.

Logical Connectors

The essays generated by ChatGPT use a variety of logical connectors to compare and contrast different perspectives, add supporting points to an argument, show cause and effect relationships and summarize and conclude the discussion. These connectors help to create a coherent and logically structured essay, guiding the reader through the arguments and supporting a clear flow of ideas.

To begin with, an essay discussing environmental conservation might use transitional phrases such as “Firstly,” “Additionally,” and “In conclusion” to guide the reader through the argument. While this ensures clarity, the lack of varied and more sophisticated transitional techniques can make the essay seem mechanical.

In terms of the logical connectors, they are used throughout the essay to enhance clarity and ensure a smooth flow between sentences and paragraphs. Each connector logically links ideas, guiding the reader through the argument seamlessly.

The use of “while” effectively contrasts the current efforts with the need for more measures, highlighting the essay’s main argument. This contrast sets up a clear problem-solution structure. Words like “additionally,” “also,” and “moreover” are used to expand on the points made, providing a comprehensive list of measures and reinforcing the argument for taking further action.

The frequent use of cause-and-effect connectors such as “leading to,” “thereby,” and “has led to” clarifies the relationships between actions and their outcomes, strengthening the logical progression of the argument.

The conclusion uses connectors effectively to summarize the discussion and restate the main points, ensuring that the essay ends with a clear and compelling call to action. Phrases like “It is undeniable that,” “One of the most effective ways,” and “In conclusion” help emphasize key points and signal the structure of the essays, aiding reader comprehension.

Overall, the essays demonstrate a strong command of logical connectors, using them effectively to create a cohesive, well-structured, and logically flowing argument. The connectors are varied and appropriately used, enhancing the overall readability and persuasiveness of the essays.

Textual Coherence

One of the strongest aspects of the essays was their coherence and cohesion.

In the Introduction, the essays start with a clear statement of the topic, followed by outlining the two sides of the argument. The purpose of each essay is stated at the end of the introduction, which helps in setting the context.

In the Body Paragraph 1, this paragraph discusses the advantages of staying with the same organization. The ideas are connected well, with each sentence building on the previous one. The use of transition words like “Moreover” helps in maintaining the flow.

The Body Paragraph 2 addresses the disadvantages of staying with one organization. The use of “However” signals a contrast to the previous paragraph, maintaining coherence. The sentences are logically connected, discussing the risks and potential drawbacks.

In the Conclusion, it restates the author’s opinion, summarizing the key points discussed in the essay. The use of “To conclude” indicates the end of the discussion, and the final sentence acknowledges individual preferences, adding a balanced perspective.

Lexical Resource

The lexical resource in the AI-generated essays showed limitations. They often relied on repetitive vocabulary and lacked diversity in word choice. While basic vocabulary was used accurately, the essays struggled with offering varied and nuanced expressions. This suggests that AI may not yet fully capture the richness and subtlety of language required for higher proficiency levels.

Lexical Accuracy

The essays demonstrated a solid understanding of basic vocabulary and were free from significant lexical errors. For example, in “It is undeniable that car ownership has increased rapidly over the past three decades, leading to cities becoming increasingly congested. With more people owning cars, roads have become overcrowded, thereby increasing emissions, and creating health hazards in cities”, the word “congested” is used correctly as an adjective standing after “become”.

Lexical Diversity

The range of vocabulary was narrow, with frequent repetition of words and phrases, leading to a lack of lexical richness and sophistication.

In an essay about the effects of globalization, ChatGPT might repeatedly use words like “globalization,” “economy,” and “culture” without introducing synonyms or related terms like “international integration,” “economic systems,” or “cultural exchange,” which would enrich the text.

Collocations

Common collocations are used throughout the essays written by ChatGPT such as to play/ have a(n) important/ key/ vital/ crucial role in (doing) something, or to be key factors influencing something or to meet the need of/ demand.

Academic Vocabulary

The use of vocabulary in the essays was found to be adequate but limited in variety such as firstly, secondly, Importantly; significantly; notably; another key point.

Grammatical Range and Accuracy

The AI-generated essays demonstrated strong grammatical accuracy. They exhibited proficiency in using a variety of grammatical structures correctly. AI reliably produced grammatically sound sentences, showcasing its ability to assist in teaching fundamental grammar rules. However, the essays occasionally lacked complexity in sentence structure, relying on simpler constructions.

Range of Structures

Despite the accuracy, the range of grammatical structures was limited. The essays often relied on simple and repetitive sentence patterns, lacking complex structures that would demonstrate a higher level of grammatical competence.

In an essay on the advantages of renewable energy, ChatGPT might use straightforward sentences like “Solar power is beneficial. It is renewable and clean.” To demonstrate a wider grammatical range, sentences could

be combined and varied: “Solar power, being both renewable and clean, offers numerous benefits.”

Grammatical accuracy

As can be seen from the analysis, the grammatical structures used in the essays generated by ChatGPT were correct and appropriate, contributing to the clarity of the essays.

Some evidence can be illustrated in “It helps employees to build a range of adaptable skills that come in handy in later stages of their careers. Switching organizations also exposes individuals to new technology and methods, and helps to keep them updated and relevant for the industry.”

Mechanics

The findings show that good uses of mechanics are seen throughout the essays such as full stops, commas, semi-colons and the like. There are no mechanic mistakes found in the essays written by ChatGPT.

It can be illustrated as “Furthermore, the extension of life expectancy can lead to greater emotional fulfillment, given that people are often able to spend more time with their loved ones, and foster more meaningful relationships with family and friends.”

Complexity and Clarity

As can be observed from the study’s findings, the essays generated by ChatGPT show a good command of grammar, with accurate and correct usage throughout.

The essays contain complex structures; for example, adverbial clauses (while) are used “While most cities have implemented policies to reduce such problems, there is a growing need for more active and far-reaching measures to curb car usage. This essay will explore the extent to which cities have become one big traffic jam as a result of increased car ownership, and discuss several measures that governments can take to discourage car usage.”.

FINDINGS AND DISCUSSION

The study highlights that while ChatGPT-generated essays had strengths in coherence, cohesion, and grammatical accuracy, which is in alignment with what Banchs and Li (2012) found, they were limited in lexical variety and grammatical range. Additionally, the essays lacked originality and critical thinking, which are crucial for higher band scores in IELTS writing. In other words, AI-generated essays excel in certain aspects such as coherence and grammatical accuracy, they fall short in lexical diversity and originality.

The findings suggest that AI can effectively support language learners in mastering basic writing skills but may require augmentation with human instruction to develop higher-order skills such as critical thinking and creative expression. ChatGPT essays lacked a personal touch and came across as robotic, so the essays did not have a clear argument, which made them less convincing. As can be seen in Appendix 3, the drawbacks are avoided in the human-written essay where the originality and clear arguments are given, showing the writing identity of the writer. These findings are consistent with previous studies that have highlighted the limitations of AI-generated essays (Alsallal, 2016; Amini & Susanti, 2024; Cocarascu & Toni, 2018).

Although ChatGPT essays can generate coherent and fluent text, they lack the critical thinking required for effective writing. The results showed that ChatGPT can generate reliable and valid scores when assessed against specific criteria. The ChatGPT system's performance was compared to that of human markers, and the scores were highly correlated. However, the ChatGPT system falls short in providing comprehensive feedback (Barrot, 2023; Ray, 2023). Participants appreciated the human aspects of the feedback, such as the feedback's personalization and critical thinking, which is opposite to the result of Lametti (2022).

As stated before, several studies have concluded that ChatGPT can be useful in assessing students' written work for specific criteria, such as grammar, syntax, and spelling accuracy in the educational context (Banchs & Li, 2012; Xiao et al., 2020). Students can use ChatGPT to assess their self-pace of writing development by asking ChatGPT to give suggestions to their writing work.

The findings suggest that while AI can serve as a valuable tool in language education for basic writing skills, its current limitations necessitate caution in its use as a sole means of instruction. Integrating AI tools alongside traditional teaching methods may enhance learning outcomes by focusing on complementary strengths. Results show that ChatGPT can produce high-quality essays, but there is still room for improvement.

RECOMMENDATIONS

Based on the results of our study, we recommend that ChatGPT essays should be used as a tool for assistance and not as a replacement for human writing. While ChatGPT essays can be helpful in generating ideas and organizing thoughts, they should not be used as a substitute for writing skills. It is important to note that effective writing requires critical thinking, originality, and a personal touch, which AI-generated text cannot replicate.

The increasing use of ChatGPT in educational settings raises concerns about their effectiveness in evaluating students' written work. This research aims to evaluate essays written by ChatGPT, comparing their performance with those of human markers. Through a qualitative approach that includes a questionnaire and interviews with participants, this research explores the reliability and validity of the assessment process using ChatGPT. The results show that ChatGPT has a high level of reliability and validity when evaluated for specific criteria. However, they lack the human perspectives, critical thinking, and cultural and social sensitivity that human markers bring, which limit their ability to

provide comprehensive feedback to the students. Based on the findings, it is recommended that ChatGPT be used in conjunction with human markers to evaluate essays. This approach provides a balance between the speed and efficiency of ChatGPT and the personalized and insightful feedback provided by human markers. The study's findings also underscore the importance of training and educating human markers on how to use ChatGPT systems to ensure they are used optimally. With the evolution of technology, artificial intelligence (AI) has emerged as a paradigm shift to solve complex problems. ChatGPT is an innovative model of AI that allows individuals to communicate in natural language.

Recommendations for Teachers

Educators should utilize AI-generated essays as supplementary resources to reinforce grammar and coherence skills. Emphasizing critical thinking and creativity through human-led instruction remains crucial. First, they can use ChatGPT essays as examples to illustrate proper structure and grammatical accuracy. Second, they can encourage students to identify areas where the AI-generated essays fall short, such as lexical variety and critical analysis, to enhance their own writing skills. Third, teachers can use ChatGPT-generated essays to demonstrate proper essay structure and grammatical accuracy to students. Fourth, they can encourage students to critique ChatGPT essays, identifying areas where more lexical variety or deeper analysis could improve the writing. Teacher can use ChatGPT to provide feedback or supplement lexical exercises for students.

Recommendations for Students

Students can benefit from using AI tools for practice and immediate feedback on structural and grammatical aspects of writing. However, they should also engage in activities that foster original thought and diverse vocabulary usage. First, students can utilize ChatGPT for brainstorming and generating ideas, but focus on adding personal insights and diverse vocabulary to improve the quality of their essays.

Second, they can practice writing complex sentences and using varied grammatical structures to develop a more sophisticated writing style. Third, they can use ChatGPT to generate ideas and outlines for essays, but avoid relying solely on the AI-generated text. Fourth, they can focus on adding personal insights, diverse vocabulary, and varied grammatical structures to their own writing to make it more sophisticated. Finally, students can regularly practice writing complex sentences and using a wider range of vocabulary to improve overall writing skills by using the Chat to upgrade their vocabulary size. The findings suggest that ChatGPT can be a useful tool for assistance in language learning, particularly in demonstrating coherent structure and correct grammar. However, it should not replace human creativity and critical thinking in essay writing.

LIMITATIONS

This study has several limitations. First, the relatively small sample size restricts the generalizability of the findings, making it difficult to draw broader conclusions about AI-generated essays. Second, the study focused on specific aspects of essay quality, leaving other important dimensions — such as cultural and social sensitivity — unexamined. Third, the evaluation relied on human judgment, which, despite efforts to maintain consistency, remains inherently subjective. Additionally, the study assessed essays generated by a single version of ChatGPT, and future research could explore variations across different AI models. To gain deeper insights, questionnaires and interviews were conducted, but further studies could adopt more comprehensive evaluation frameworks, including the latest AI models, to enhance the analysis of essay quality.

CONCLUSION

In conclusion, this study evaluated the essays written by ChatGPT, an AI language model designed to mimic human conversation. While ChatGPT

essays showed positive attributes such as coherence and fluency, they lacked originality and critical thinking. We recommend that ChatGPT essays should be used as a tool for assistance and not as a replacement for human writing. It is important to note that effective writing requires critical thinking, originality, and a personal touch, which AI-generated text cannot replicate. The study's findings emphasize the importance of using ChatGPT in conjunction with human markers to provide students with insightful and personalized feedback.

Overall, AI technologies present promising opportunities for enhancing language education, particularly in writing skills development. While AI-generated essays exhibit strengths in certain areas, they currently lack the sophistication required for nuanced and original expression. Future advancements in AI, coupled with thoughtful integration into pedagogical practices, hold potential for transformative educational outcomes in language learning.

REFERENCES

- Alsallal, M. (2016). A machine learning approach for plagiarism detection (Doctoral dissertation, Coventry University).
- Alsaweed, W., & Aljebreen, S. (2024). Investigating the Accuracy of ChatGPT as a Writing Error Correction Tool. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, 14(1), 1-18.
- Amini, F., & Susanti, A. (2024). ChatGPT: Enhancing Students' Writing Skills For EFL Students In Descriptive Text. *International Journal of Research and Innovation in Social Science*, 8(10), 2273-2285.
- Banchs, R. E., & Li, H. (2012, July). IRIS: a chat-oriented dialogue system based on the vector space model. *In Proceedings of the ACL 2012 System Demonstrations* (pp. 37-42).

- Barrot, J. S. (2023). *Using ChatGPT for second language writing: Pitfalls and potentials*. *Assessing Writing*, 57, 100745.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877-1901.
- Cocarascu, O., & Toni, F. (2018). Combining deep learning and argumentative reasoning for the analysis of social media textual content using small data sets. *Computational Linguistics*, 44(4), 833-858.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37-46.
- Cohen, L., Manion, L., & Morrison, K. (2017). Validity and reliability. In *Research methods in education* (pp. 245-284). Routledge.
- Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological testing and assessment: An introduction to tests and measurement*. Mayfield Publishing Co.
- de Rivero, M., Tirado, C., & Ugarte, W. (2023). Formalstyler: Gpt-based model for formal style transfer with meaning preservation. *SN Computer Science*, 4(6), 739.
- Ghafar, Z. N. (2023). ChatGPT: a new tool to improve teaching and evaluation of second and foreign languages a review of ChatGPT: the future of education. *International journal of applied research and sustainable sciences*, 1(2), 73-86.
- IELTS. (2019). *IELTS writing band descriptors (public version)*. <https://www.ielts.org/-/media/pdfs/writing-band-descriptors-task-1-task-2.ashx>
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, 103, 102274.

- Lametti, D. (2022). *AI could be great for college essays*. Slate. Last accessed December, 15, 2022.
- Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121-154.
- Sallam, M. (2023, March). ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *In Healthcare (Vol. 11, No. 6, p. 887)*. MDPI.
- Taecharungroj, V. (2023). “What can ChatGPT do?” Analyzing early reactions to the innovative AI chatbot on Twitter. *Big Data and Cognitive Computing*, 7(1), 35.
- Wilson, J., & Roscoe, R. D. (2020). Automated writing evaluation and feedback: Multiple metrics of efficacy. *Journal of Educational Computing Research*, 58(1), 87-125.
- Xiao, R., Guo, W., Zhang, Y., Ma, X., & Jiang, J. (2020, December). Machine learning-based automated essay scoring system for Chinese proficiency test (HSK). *In Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval* (pp. 18-23).

APPENDICES

Appendix 1: Essay topics

Prompt 1

Some people work for the same organization all their working life. Others think it is better to work for different organizations. Discuss both these views and give your own opinion. You should give reasons for your answer and include any relevant examples from your own

knowledge or experience. You should write at least 250 words. You should spend about 40 minutes on this task.

Prompt 2

One of the consequences of improved medical care is that people are living longer and life expectancy is increasing. Do you think the advantages of this outweigh the disadvantages? You should give reasons for your answer and include any relevant examples from your own knowledge or experience. You should write at least 250 words. You should spend about 40 minutes on this task.

Prompt 3

Write a well-developed essay on the following topic: “Car ownership has increased so rapidly over the past thirty years that many cities in the world are now ‘one big traffic jam.’ How true do you think this statement is? What measures can governments take to discourage people from using their cars?” You should give reasons for your answer and include any relevant examples from your own knowledge or experience. You should write at least 250 words. You should spend about 40 minutes on this task.

Prompt 4

Write a well-developed essay on the following topic: “In some countries an increasing number of people are suffering from health problems as a result of eating too much fast food. It is therefore necessary for governments to impose a higher tax on this kind of food. To what extent do you agree or disagree with this opinion?” You should give reasons for your answer and include any relevant examples from your own knowledge or experience. You should write at least 250 words. You should spend about 40 minutes on this task.

Prompt 5

Write a well-developed essay on the following topic: “Nowadays technology is increasingly being used to monitor what people are

saying and doing (for example, through cell phone tracking and security cameras). In many cases, the people being monitored are unaware that this is happening. Do you think the advantages of this development outweigh the disadvantages?” You should give reasons for your answer and include any relevant examples from your own knowledge or experience. You should write at least 250 words. You should spend about 40 minutes on this task.

Prompt 6

Write a well-developed essay on the following topic: “In the past, when students did a university degree, they tended to study in their own country. Nowadays, they have more opportunity to study abroad. What are the advantages and disadvantages of this development?” You should give reasons for your answer and include any relevant examples from your own knowledge or experience. You should write at least 250 words. You should spend about 40 minutes on this task.

Prompt 7

Write a well-developed essay on the following topic: “The continued rise in the world’s population is the greatest problem faced by humanity at the present time. What are the causes of this continued rise? Do you agree that it is the greatest problem faced by humanity?” You should give reasons for your answer and include any relevant examples from your own knowledge or experience. You should write at least 250 words. You should spend about 40 minutes on this task.

Prompt 8

Write a well-developed essay on the following topic: “It is sometimes said that people should be encouraged to get married before they turn 30, as this is best both for the individual and for the society. Do you agree or disagree?” You should give reasons for your answer and include any relevant examples from your own knowledge or experience. You should write at least 250 words. You should spend about 40 minutes on this task.

Prompt 9

Write a well-developed essay on the following topic: “Some people get into debt by buying things they don’t need and can’t afford. What are the reasons for this behavior? What action can be taken to prevent people from having this problem?” You should give reasons for your answer and include any relevant examples from your own knowledge or experience. You should write at least 250 words. You should spend about 40 minutes on this task.

Prompt 10

Write a well-developed essay on the following topic: “Many people like to wear fashionable clothes. Why do you think this is the case? Is this a good thing or a bad thing?” You should give reasons for your answer and include any relevant examples from your own knowledge or experience. You should write at least 250 words. You should spend about 40 minutes on this task.

Appendix 2: Band scoredescriptors

IELTS

WRITING TASK 2: Band Descriptors (public version)

Band	Task response	Coherence and cohesion	Lexical resource	Grammatical range and accuracy
9	<ul style="list-style-type: none"> fully addresses all parts of the task presents a fully developed position in answer to the question with relevant, fully extended and well-supported ideas 	<ul style="list-style-type: none"> uses cohesion in such a way that it attracts no attention skillfully manages paragraphing 	<ul style="list-style-type: none"> uses a wide range of vocabulary with very natural and sophisticated control of lexical features, rare minor errors occur only as ‘slips’ 	<ul style="list-style-type: none"> uses a wide range of structures with full flexibility and accuracy, rare minor errors occur only as ‘slips’
8	<ul style="list-style-type: none"> sufficiently addresses all parts of the task presents a well-developed response to the question with relevant, extended and supported ideas 	<ul style="list-style-type: none"> sequences information and ideas logically manages all aspects of cohesion well uses paragraphing sufficiently and appropriately 	<ul style="list-style-type: none"> uses a wide range of vocabulary fluently and flexibly to convey precise meanings skillfully uses uncommon lexical items but there may be occasional inaccuracies in word choice and collocation produces rare errors in spelling and/or word formation 	<ul style="list-style-type: none"> uses a wide range of structures the majority of sentences are error-free makes only very occasional errors or inappropriacies
7	<ul style="list-style-type: none"> addresses all parts of the task presents a clear position throughout the response extends and supports main ideas, but there may be a tendency to over-generalise and/or supporting ideas may lack focus 	<ul style="list-style-type: none"> logically organises information and ideas; there is clear progression throughout uses a range of cohesive devices appropriately although there may be some under-use presents a clear central topic within each paragraph 	<ul style="list-style-type: none"> uses a sufficient range of vocabulary to allow some flexibility and precision uses less common lexical items with some awareness of style and collocation may produce occasional errors in word choice, spelling and/or word formation 	<ul style="list-style-type: none"> uses a variety of complex structures produces frequent error-free sentences has good control of grammar and punctuation but may make a few errors
6	<ul style="list-style-type: none"> addresses all parts of the task although some parts may be more fully covered than others presents a relevant position although the conclusions may become unclear or repetitive presents relevant main ideas but some may be inadequately developed/unclear 	<ul style="list-style-type: none"> arranges information and ideas coherently and there is a clear overall progression uses cohesive devices effectively, but cohesion within and/or between sentences may be faulty or mechanical may not always use referencing clearly or appropriately uses paragraphing, but not always logically 	<ul style="list-style-type: none"> uses an adequate range of vocabulary for the task attempts to use less common vocabulary but with some inaccuracy makes some errors in spelling and/or word formation, but they do not impede communication 	<ul style="list-style-type: none"> uses a mix of simple and complex sentence forms makes some errors in grammar and punctuation but they rarely reduce communication
5	<ul style="list-style-type: none"> addresses the task only partially, the format may be inappropriate in places expresses a position but the development is not always clear and there may be no conclusions drawn presents some main ideas but these are limited and not sufficiently developed; there may be irrelevant detail 	<ul style="list-style-type: none"> presents information with some organisation but there may be a lack of overall progression makes inadequate, inaccurate or over-use of cohesive devices may be repetitive because of lack of referencing and substitution may not write in paragraphs, or paragraphing may be inadequate 	<ul style="list-style-type: none"> uses a limited range of vocabulary, but this is minimally adequate for the task may make noticeable errors in spelling and/or word formation that may cause some difficulty for the reader 	<ul style="list-style-type: none"> uses only a limited range of structures attempts complex sentences but these tend to be less accurate than simple sentences may make frequent grammatical errors and punctuation may be faulty; errors can cause some difficulty for the reader
4	<ul style="list-style-type: none"> responds to the task only in a minimal way or the answer is tangential, the format may be inappropriate presents a position but this is unclear presents some main ideas but these are difficult to identify and may be repetitive, irrelevant or not well supported 	<ul style="list-style-type: none"> presents information and ideas but these are not arranged coherently and there is no clear progression in the response uses some basic cohesive devices but these may be inaccurate or repetitive may not write in paragraphs or their use may be confusing 	<ul style="list-style-type: none"> uses only basic vocabulary which may be used repetitively or which may be inappropriate for the task has limited control of word formation and/or spelling; errors may cause strain for the reader 	<ul style="list-style-type: none"> uses only a very limited range of structures with only rare use of subordinate clauses some structures are accurate but errors predominate, and punctuation is often faulty
3	<ul style="list-style-type: none"> does not adequately address any part of the task does not express a clear position presents few ideas, which are largely undeveloped or irrelevant 	<ul style="list-style-type: none"> does not organise ideas logically may use a very limited range of cohesive devices, and those used may not indicate a logical relationship between ideas 	<ul style="list-style-type: none"> uses only a very limited range of words and expressions with very limited control of word formation and/or spelling; errors may severely distort the message 	<ul style="list-style-type: none"> attempts sentence forms but errors in grammar and punctuation predominate and distort the meaning
2	<ul style="list-style-type: none"> barely responds to the task does not express a position may attempt to present one or two ideas but there is no development 	<ul style="list-style-type: none"> has very little control of organisational features 	<ul style="list-style-type: none"> uses an extremely limited range of vocabulary, essentially no control of word formation and/or spelling 	<ul style="list-style-type: none"> cannot use sentence forms except in memorised phrases
1	<ul style="list-style-type: none"> answer is completely unrelated to the task 	<ul style="list-style-type: none"> fails to communicate any message 	<ul style="list-style-type: none"> can only use a few isolated words 	<ul style="list-style-type: none"> cannot use sentence forms at all
0	<ul style="list-style-type: none"> does not attempt does not attempt the task in any way writes a totally memorised response 			

Appendix 3: Sample from ChatGPT essay and human-written essay

Prompt 1

ChatGPT-written

The question of whether it is better to remain with one organization throughout an entire career or to work for different organizations has been a subject of debate for some time. Those who work for the same organization throughout their career often develop an intimate knowledge of the company's culture, practices, and products. By contrast, those who work for different organizations have a more diverse range of experiences, often spanning different industries and markets. In this essay, I will discuss both sides of the argument, and offer my own opinion on the matter.

One of the primary advantages of working for the same organization throughout one's career is the sense of loyalty and dedication that can be established towards a company. With time and experience, an employee in such an organization becomes more familiar with the company's demands, procedures, and culture. This, in turn, can help the employee make better decisions regarding promotions or work goals. Moreover, a long-standing relationship with a company can help people feel more secure in their job, which can lead to a sense of job satisfaction.

However, there are also some drawbacks to staying with one organization throughout your career. One of the biggest disadvantages is the possibility of stagnation, where one may not grow or develop new skills for the industry. Also, the company may experience inevitable changes, and if the employee has not kept up with new trends, their skills may become outdated. Thus, there are risks associated with limiting one's potential for career growth and development.

On the other hand, working for different organizations in one's career can offer employees more diverse experiences and insights. It enables employees to explore a variety of roles and positions across different industries, which can help in developing a broadened perspective about the professional world. It helps employees to build a range of adaptable skills that come in handy in later stages of their careers. Switching

organizations also exposes individuals to new technology and methods, and helps to keep them updated and relevant for the industry.

To conclude, my opinion is that working for different organizations throughout one's career is a better option. Switching jobs and industries allows people to gain new experiences, insights and connections in various areas that can widen their professional network. It allows individuals to expand their knowledge base and gain new skills that can be transferred to other organizations. Ultimately, the decision between the two options is up to the individual and depends on factors such as their career goals, work experience, and personal preferences.

Human-written

Today, different individuals have a varied opinion about the work options. Some people prefer working for same company throughout their life, while others believe in switching their companies. Both viewpoints will be critiqued before reaching a reasoned conclusion.

On the one hand, there are a group of individuals who dedicate their professional life for one corporation only. The primary reason for following this is the stability which can be achieved only by working at the same place for a longer time. This will not only allow employees to gain thorough knowledge about the application but also increase the chances of recognitions and promotions. Thus, it is obvious that professional life will be at the advantage in case he is employed in the same industry for a lifetime.

On the other hand, there are some individuals who believe that they are benefited by altering companies. This is because such folks can get exposure to different working environments, thereby, allowing them to get trained about different systems available in the market. This can increase the market value of an individual, resulting in receiving offers with a higher salary, and hence, financial stability can be attained at an earlier stage. In conclusion, this will enable individual to lead a successful and secure personal life.

Upon the analysis of both the views, it is apparent that varying jobs should be preferred over other since it can provide better family life. Thus, it is recommended that public will keep themselves motivated for shifting industries which can help them to enjoy their life and fulfil family requirements.