

# The Effectiveness of Automated Written Corrective Feedback on L2 Learners' Revision Outcomes: A Case for ChatGPT

Yushan Zhou<sup>1,a,\*</sup>

<sup>1</sup>*School of Humanities and Social Science, The Chinese University of Hong Kong, Shenzhen, Shenzhen, China*

<sup>a</sup>*120030091@link.cuhk.edu.cn*

*\*Corresponding author*

**Abstract:** *Technological advancements have significantly reshaped L2 teaching pedagogy, leading to extensive application of automated written evaluation (AWE) tools in L2 learning. While previous research has mainly focused on lower-order feedback from traditional AWE tools, higher-order issues have been largely overlooked. The emergence of ChatGPT, a burgeoning large language model, introduces fresh perspectives on automated written corrective feedback (AWCF). This study investigates ChatGPT's competence in tackling diverse writing challenges, including both lower-order and higher-order aspects, to assist L2 learners in their IELTS writing. It further identifies the comparative strengths and shortcomings of ChatGPT against traditional AWE tools. The findings underscore the importance of tailoring ChatGPT's feedback to individual needs, highlighting its potential as an AWE resource for writing support. These insights inform the incorporation of ChatGPT in L2 writing instruction, enhancing language proficiency and assessment methods in education.*

**Keywords:** *Second Language Acquisition; Automated Written Corrective Feedback; ChatGPT; Academic Writing*

## 1. Introduction

Automated writing evaluation (AWE) tools, such as Grammarly and Pigai, have increasingly become integral components in L2 pedagogy. These platforms have demonstrated capability in enhancing L2 writing quality, complementing traditional teacher feedback <sup>[1]</sup>. Nonetheless, despite their efficacy, these established traditional AWE systems often struggle to address intricate writing errors, especially those related to logical progression. Consequently, they may produce generic feedback that does not fully address diverse learning needs <sup>[2]</sup> <sup>[3]</sup>. Although there is an evident focus on identifying AWE tools that provide personalized feedback <sup>[4]</sup>, much of the recent research is predominantly concerned with how AWE tools address lower-order writing. Empirical investigations into ChatGPT's potential are scant, with discussions largely centered on anecdotal experiences. This gap in previous studies necessitates an empirical exploration of ChatGPT's role in the scope of L2 pedagogy, with a particular emphasis on its effectiveness as an AWE tool.

Chat Generative Pre-Trained Transformer (abbreviated as ChatGPT), a cutting-edge language model, emerges as a promising contender in this regard. Unlike traditional AWE tools, ChatGPT can offer personalized feedback through iterative chatbot-human interactions, addressing not only foundational grammatical errors but also delving into advanced facets such as content organization, register appropriateness, and logical coherence <sup>[5]</sup>. By remedying the insufficiency of traditional AWE tools, particularly contextual understanding, ChatGPT improves the quality of automated written corrective feedback (AWCF).

This qualitative study seeks to assess ChatGPT's effectiveness in aiding students with both fundamental and advanced challenges in IELTS academic writing tasks. It also aims to capture students' perceptions concerning the feedback from this novel tool. The results intend to enrich the burgeoning discourse on the integration of state-of-the-art language models, such as ChatGPT, in the academic writing milieu.

## 2. Literature Review

### 2.1 Overview of Feedback and Traditional AWE Tools

Feedback is crucial in language learning, revealing the discrepancies between a learner's performance and the targeted outcomes [6]. With the rise of technology-driven feedback in language instruction, traditional AWE tools have gained prominence in second language acquisition (SLA), particularly for their ability in error detection and correction [5][7][8][9]. Notably, these tools adeptly address both "lower-order" concerns, centered on specific grammar issues, and "higher-order" aspects, which emphasize subjective components like thesis and organization [10]. These components of writing are deeply intertwined with critical thinking and innovative reasoning [11]. Therefore, leveraging AWE tools in emphasizing both writing dimensions is vital in SLA.

### 2.2 Traditional AWE in Addressing Lower and Higher-order Issues

Lower-order issues, which include capitalization, grammar, spelling, and punctuation (often referred to as CGSP), critically influence the quality and clarity of L2 academic writing. In contrast, higher-order challenges pertain to interpretation, analysis, and structuring, evident in the formation of a coherent thesis, understanding the audience, logical organization, and compelling argumentation [12]. In addressing these challenges, AWE tools such as Pigai, Criterion, and Grammarly have gained traction. Pigai emphasizes continuous practice and improvement [13], Criterion focuses on linguistic features with in-depth error reports [14], and Grammarly offers a wide-ranging toolset aimed at enhancing clarity and precision [15].

Yet, these tools exhibit limitations. John and Wool [9] observed that they address lower-order issues only to a limited extent (<50%) and often overlook mistakes that human reviewers can identify [16]. Many studies have indicated that while students rely on traditional AWE tools for basic errors, these platforms often struggle with higher-level tasks like content creation [14]. Consequently, emerging technologies like ChatGPT, capable of providing detailed feedback on intricate writing challenges, are drawing increased interest.

### 2.3 The Application of ChatGPT in Educational Settings

Advancements in AI have led to the emergence of numerous English learning applications pivotal in SLA. Notably, ChatGPT demonstrates proficiency in diverse language tasks, producing human-like responses, and signaling its academic potential [17]. As a leading large language model, ChatGPT facilitates reading, writing, critical thinking, research, and problem-solving, positioning itself as an essential asset for academic growth [5].

#### 2.3.1 Potential of ChatGPT in Addressing Lower and Higher-order Issues

In the realm of academic writing, ChatGPT, as an AWE tool, has showcased significant prowess. Beyond generating cohesive text and providing grammar and style recommendations, it assists in formulating well-structured arguments, proving invaluable to both students and scholars [5][18][19]. Zohery [20] highlighted its proficiency in addressing lower-order issues such as vocabulary refinement, grammar checking, and spelling corrections. Complementing these foundational capabilities, ChatGPT delivers customized feedback, proposing resources that resonate with learners' objectives [5][21].

ChatGPT also stands out on the front of higher-order issues. It aids in crafting content that aligns with academic benchmarks, ensuring clarity and conformity to recognized writing conventions [7]. Regardless of the document type – be it an essay, report, or research paper – ChatGPT maintains uniformity in grammar, style, and voice, imperative for upholding its integrity [20]. The model identifies discrepancies in expansive texts, ensuring textual cohesion. Furthermore, bolstered by its nuanced understanding, ChatGPT appraises the logical progression, assesses the strength of arguments, and introduces refinements. Such insightful feedback acquaints L2 learners with established academic writing criteria and cultivates analytical reasoning [5][22].

### 2.4 Prompts in ChatGPT

ChatGPT's potential to deliver individualized learning experiences stems from its ability to furnish automated, corrective feedback aligned with users' distinct writing needs and learning preferences [5]. This tool generates contextually relevant continuations based on the conversation's "prompt" and its intrinsic language understanding. The efficacy of feedback provision via ChatGPT largely hinges on the

prompts given [23]. Therefore, adhering to specific guiding principles becomes imperative to obtain optimal and pertinent feedback [7]. Such principles include choosing words carefully, defining the purpose and focus of the conversation, being specific and concise, providing context, encouraging expansion on responses, and utilizing the "stop" command when needed. By crafting effective prompts consistent with the principles as outlined, learners can fine-tune the feedback received, ensuring the corrective feedback provided aligns with their unique writing needs and pedagogical objectives. In this light, L2 learners can more effectively harness the potential of ChatGPT.

### 2.5 The Present Study

While traditional AWE tools can address basic writing challenges, they often fail to provide accurate, comprehensive, and higher-order writing suggestions. ChatGPT, with its advanced language comprehension, presents a promising alternative to address these limitations. This research adopts a pedagogical approach, combined with textual and thematic analysis, to assess ChatGPT's feedback on university students' IELTS academic writing. The study focuses on two primary research questions.

RQ1: To what extent does ChatGPT surpass traditional models in providing automated corrective feedback?

RQ2: How well can ChatGPT assist in addressing lower and higher-order issues respectively in IELTS writing?

## 3. Methodology

### 3.1 Participants

This study utilized a convenience sampling method, selecting participants from undergraduate students at The Chinese University of Hong Kong, Shenzhen (CUHKSZ). The study comprised six participants (see Table 1 for more information).

Table 1 Participant information

Name	Gender	Grade	Major	English Proficiency (Test Level)	IELTS Experience
Annie	Female	Y2	Translation	Passed TEM-4	Not Taken
Shirley	Female	Y3	English Studies	Passed TEM-4	Taken (6.5)
Yuki	Female	Y3	Translation	Passed TEM-4	Taken (6.5)
Lan	Female	Y2	Translation	Passed TEM-4	Taken (6)
Wendy	Female	Y4	Finance	Passed CET-6	Taken (6.5)
Leo	Male	Y3	Computer Science	Passed CET-6	Not Taken

### 3.2 Data Collection and Analysis

**Stage 1 Workshop:** Participants received training on IELTS Task 2, focusing on its complex logical structures, precise grading descriptors, and optimal prompting techniques with ChatGPT. They were then required to write an argumentative paragraph on a IELTS Task 2 topic concerning global rural-urban migration instructed as follows.

*In many countries around the world, rural people are moving to cities, so the population in the countryside is decreasing. Do you think this is a positive or a negative development?*

During the initial drafting, they abstained from using any AI-enhanced or AWE tools. After the initial draft, they engaged with ChatGPT for feedback and revised their drafts accordingly.

**Stage 2 One-on-one Interviews:** Post-workshop, individual interviews delved into participants' experiences and perceptions regarding ChatGPT. During these interviews, questions related to three overarching themes were discussed: 1) evaluation and enhancement of AWE Tools, including general evaluation of AWE tools and potential enhancements for AI tools, 2) ChatGPT's role in Writing Support, specifically how ChatGPT aids in IELTS, and 3) user perspectives on AI-Assisted Writing. To ensure participants' comfort in expressing their views, interviews were conducted in Chinese, their L1, and

were recorded upon participants' consent.

In this research, the effectiveness of corrective feedback offered by ChatGPT is scrutinized through participants' academic IELTS writings based on four official grading descriptors: grammatical range and accuracy, lexical resource (lower-order criteria), and task response, along with coherence and cohesion (higher-order criteria). The objective is to evaluate ChatGPT's performance as an AWE tool in aiding participants' IELTS writing revisions. The initial drafts (draft 1) and revised revisions (draft 2) were assessed based on the four primary assessment criteria for IELTS writing: task response (TR), coherence and cohesion (CC), lexical resource (LR), and grammatical range and accuracy (GRA). The TR criterion measures the degree to which participants understood the task and the relevance of information. The CC criterion assesses the logical arrangement of ideas and how effectively transitions are used to maintain continuity. On the other hand, the LR criterion evaluates the precision, diversity, and idiomatic usage of the vocabulary within the drafts. Finally, the GRA criterion examines the sophistication of sentence structures and the accuracy of grammar employed, pinpointing any rectifications made between the two drafts.

The interview transcripts underwent thematic analysis to discern patterns and insights. To minimize potential biases and to reinforce the study's validity, the analysis was reviewed and revised two weeks after the coding.

## 4. Results and Discussion

### 4.1 Textual Analysis of Writing Drafts

Participants' first and second drafts were collected and compared, based on all four IELTS grading criteria: TR, CC, LR, and GRA. Drafts were selected from the two participants for their typical representation of the grading descriptors.

#### 4.1.1 Annie

##### *Draft 1*

*To begin with, the negative consequences cannot be ignored. Firstly, the mass migration from rural to urban areas puts immense pressure on city infrastructure and resources. For instance, Shenzhen, a typical immigrant city, is criticized by the limited residential, medical, and educational resources. Moreover, the increasing population will result serious pollution. One example is that Chennai in India has a problem in providing adequate clean water and sanitation facilities. These issues adversely affect the quality of life in cities and the overall sustainability of urban environments.*

Draft 1 primarily addresses the impacts of rural-urban migration, emphasizing its negative consequences. Beginning with a somewhat broad and vague thesis, the essay then provides examples of Shenzhen and Chennai to demonstrate strains on city infrastructure and pollution. However, while these examples align directly with the topic, they lack in-depth elaboration. Moving from content to form, the vocabulary is on point, but there are minor vocabulary redundancies and some collocational inaccuracies. In terms of grammar, a couple of preposition misplacements and a punctuation error are evident, and the text would also benefit from more intricate sentence constructions.

##### *Draft 2*

*To commence, it is imperative to acknowledge the detrimental ramifications that arise. Primarily, the substantial influx of individuals from rural to urban locales exerts tremendous strain on the infrastructure and resources of cities. Take, for instance, Shenzhen, a prototypical city for immigrants, which faces criticism due to its scarcity of residential, medical, and educational provisions. Furthermore, the expanding populace leads to grave consequences in terms of pollution. A case in point is the predicament faced by Chennai, India, where there is a dearth of adequate clean water and sanitation facilities. These predicaments significantly impede the quality of life within urban areas and pose a considerable threat to the overall sustainability of urban environments.*

This revised version presents a more coherent argument with a more formal and academic introductory tone set with "To commence". In this draft, four explicit cohesive devices ("To commence," "Primarily" "Take, for instance," and "Furthermore") are utilized, reflecting a 33% increase from Draft 1, which aids in creating a smoother flow. Vocabulary in Draft 2 is not only richer and more advanced, with words like "detrimental ramifications" and "influx," but also showcases better collocation. The examples are described in greater detail, and the argument's progression is clearer. Grammatically, the

draft mends earlier errors, offering improved accuracy. In essence, Draft 2 excels in terms of cohesion, vocabulary breadth, and grammatical precision when juxtaposed with the first draft.

#### **4.1.2 Shirley**

##### *Draft 1*

*One of the negative consequences of this population shift is the strain it places on urban resources. As more people move into cities, there is a heightened demand for natural and industrial resources. In other words, the overcrowding can lead to increased pressure on utilities such as water, electricity, and transportation systems. In the long run, the overload of resource utilization will result in environmental damage and decreased quality of life.*

The narrative focuses on the strain on urban resources in Shirley's draft 1. When addressing task response, she covers one negative outcome: overcrowding, which lacks support for her central theme of urban strain with other perspectives. In terms of Coherence, she presents a clear thesis and follows with detailed, step-by-step explanations. The argument evolves logically, moving from the immediate repercussions of the population shift to long-term environmental damages. The draft employs various cohesive devices like "In other words" to elucidate the logical progression of ideas. While the vocabulary aptly captures the issue, phrases like "overload of resource utilization" and "heightened demand," suggest linguistic patterns influenced by Chinese, potentially compromising clarity. Grammatically, the draft maintains clarity but might benefit from varied sentence structures.

##### *Draft 2*

*One negative consequence of this population shift is the strain it places on urban resources. As more people move into cities, there is an increased demand for natural resources such as water and energy sources, as well as industrial resources. This overcrowding can lead to a heightened pressure on utilities, including water, electricity, and transportation systems. Consequently, it can result in environmental damage and a decreased quality of life. Moreover, the strain on these resources may exacerbate existing issues related to pollution and limited access to basic services. Considering these detrimental effects, the population shift from rural to urban areas should be considered a negative development.*

Draft 2 presents a more expansive argument concerning the effects of population shift. The essay's introduction is consistent with draft 1 but with more supporting perspectives on the results caused by overcrowding. Furthermore, the incorporation of pollution, limited access to service, and the concluding statement aligns perfectly with the writing's theme, presenting a comprehensive argument. To enhance overall coherence, the draft employs a wider range of cohesive devices. Comparatively, five such devices ("One negative consequence," "As more people," "This overcrowding," "Consequently," and "Moreover") in draft 2 increase by 25 %, enhancing the logical connection between ideas. Vocabulary is both diverse and precisely used in draft 2, and inaccurate collocations were corrected. Grammatically, the draft exhibits sophistication with a mix of compound and complex sentences. By contrast, Draft 2 provides a broader, more in-depth perspective than Draft 1.

In reviewing the revised drafts, it is evident that ChatGPT's influence significantly enhanced both lower and higher-order aspects. For lower-order concerns, both drafts exhibited refined vocabulary, enhanced logical flow, and reduced grammatical errors. On the higher-order front, there was a discernible improvement in detailed logical elaboration.

## **4.2 Thematic analysis of Individual Interviews**

### **4.2.1 Evaluating AWE Tools**

Participants' feedback illuminated a distinct contrast between traditional tools and ChatGPT. ChatGPT is praised for its capabilities extending beyond traditional models, particularly in the academic sector. Beyond merely offering writing feedback, it also provides research assistance, idea generation for theses, and tutoring in complex subjects to participants. By contrast, traditional models like Grammarly are perceived to be more "specialized in writing," as noted by Leo, with a more user-friendly interface. Participants unanimously expressed that while traditional tools address lexical and syntactic errors, ChatGPT facilitates broader engagement. As Lan stated, ChatGPT "doesn't just correct; it suggests and expands based on what I've written." Annie also mentioned, "It adapts to any language style based on genre and requirements."

However, alongside its recognition, participants pinpointed ChatGPT's limitations. They appreciated its comprehensive feedback, but often perceived it as AI-like, missing the nuance and intuition of human

feedback. As Shirley commented, “ChatGPT, while insightful, at times sounds too robotic.” Participants also raised concerns about the tool’s occasional propensity to provide misleading suggestions and consistency issues. In contrast, traditional AWE tools, by relying solely on a writer’s initial expression for corrections, seem to avoid machine-like, inconsistent generated content. Furthermore, when compared with traditional tools, concerns about ChatGPT’s user interface, particularly its inconvenience in the revision process, became pronounced. The sentiment often circled back to a comparison, where participants felt traditional tools, like Grammarly, felt quicker and more streamlined. Therefore, while ChatGPT is a powerful asset, it undeniably comes intertwined with its unique set of challenges.

#### ***4.2.2 ChatGPT in Addressing Lower-Order Issues***

For lower-order issues, participants generally didn’t solely rely on ChatGPT for grammar. They placed a high priority on utilizing ChatGPT to improve the precision of language and refine the writing style. Leo, for instance, always “verify the grammar with Grammarly before submitting the work”. This was echoed by others who also cross-referenced, deeming traditional AWE tools more specialized in grammar. This tendency might emerge from ChatGPT’s user interface, which lacks Grammarly’s in-text revision features. Additionally, while traditional tools enable direct text edits, ChatGPT outputs content in blocks, hampering real-time edits.

Participants also highlighted ChatGPT’s capacity to integrate more appropriate diction and sentence structures. This focus is predominantly on two key areas: First, its adeptness at tailoring vocabulary to the context is profound. It becomes evident when observing its selection of formal language for academic or business contexts and colloquial terms for daily exchanges. Lan observed that ChatGPT adopts formal terms for academic papers, while incorporating idiomatic expressions in her culture-centric genres, showcasing its context-aware capabilities. Building on this, Annie noted that ChatGPT can even tailor its language to specific styles, such as the “New Yorker’s,” depending on user preferences. Beyond mere word selection, ChatGPT is capable of curating appropriate and varied sentence structures. According to Yuki, it ensures that the sentence is not just coherent at a “micro-level” (between adjacent words) but also at a “macro-level” (across the entire writing). Wendy also stated that the model’s “capacity to balance syntactic variety with contextually appropriate diction” is pivotal in ensuring that the narrative remains diverse and coherent.

#### ***4.2.3 ChatGPT in Addressing Higher-Order Issues***

ChatGPT is recognized as surpassing traditional AWE tools in providing more comprehensive feedback on the logical development of ideas. Specifically, Yuki mentioned that ChatGPT detected logical inconsistencies in her writing. Both Wendy and Lan valued ChatGPT’s strength in enhancing the logical coherence of their writings. Wendy especially valued the tool’s suggestions for cohesive devices. Lan even mentioned utilizing ChatGPT to deepen her argumentation, enriching the content. Leo, from another perspective, pointed out a long-term potential in utilizing ChatGPT for enhancing logical thinking, but he also voiced concerns over possible over-reliance and stressed the need for critical evaluation of the feedback received.

Nevertheless, out of the six participants, four pinpointed ChatGPT’s occasional tendency to rephrase without introducing novel insights. While Annie didn’t observe a notable enhancement in logical coherence during her interaction with the tool, Shirley felt that ChatGPT marginally improved her logical flow. However, she also stated that there was no “major transformative impact,” implying that no fresh information was added. Despite these issues encountered, all participants acknowledged their lack of detailed prompts, admitting that more precise instructions could enhance ChatGPT’s performance on higher-order issues.

### ***4.3 Discussion***

This study offers insights into ChatGPT’s potential and limitations as an AWE tool in IELTS academic writing. While tools like Grammarly are perceived as adept in grammar correction, ChatGPT showcases versatility, refining diction and style for minor issues, and assisting in idea development and logical argumentation for intricate matters. Shared observations emerged from participants regarding the quality of ChatGPT’s feedback and its relation to users’ prompts. By closely analyzing user-chatbot interactions and participants’ interview transcripts, user factors were identified to influence the revision outcomes. According to the data, most participants didn’t interact deeply with ChatGPT while revising their writings, and the lack of detailed prompts provided to the AI affected its comprehension and feedback on higher-order issues. Furthermore, certain participants, at times, might not recognize the value of extending logical connections, which could intrinsically influence coherence and cohesion.

This research positions ChatGPT distinctively, given its capability of addressing both lower and higher-order concerns, highlighting its transformative potential in L2 writing pedagogy. In line with previous studies <sup>[7][17][23]</sup>, our findings reinforce ChatGPT's academic merit. A key finding of this study is the pivotal role of user prompts in the revision process, specifically in shaping the quality of feedback. This focus is vital to the research results, as it shifts the responsibility from solely the tool's effectiveness to a combined effort between the user and the AI. Furthermore, while earlier studies mainly concentrated on the effectiveness of traditional AWE tools on grammar and style corrections <sup>[5][18][20]</sup>, this research illuminates the potential of ChatGPT in addressing higher-order writing concerns, an area less explored in the current literature. However, for effective incorporation of ChatGPT in L2 learning, it's crucial to understand its strengths and limitations, ensuring that it complements, rather than replaces, traditional feedback mechanisms.

## 5. Conclusions

Upon delving into ChatGPT's capabilities, two research findings became evident. Firstly, in contrast to traditional AWE models, ChatGPT garnered favorable feedback due to its more extensive evaluations, flexibility, and aptitude in addressing complex logical concerns. Yet, challenges such as prompt misinterpretation and inconsistent feedback highlight areas for refinement for both ChatGPT and users. Secondly, ChatGPT surpassed other AWE tools in resolving IELTS writing's lower-order issues. Its proficiency in managing higher-order issues is enhanced by well-crafted prompts, emphasizing the importance of skilled user interaction and prompt design to optimize ChatGPT's potential in academic assessments. This study, however, has its limitations, including its qualitative nature, limited participant size, and IELTS-centric focus, potentially restricting its generalizability. Comprehensive research across various writing genres is essential to ascertain ChatGPT's effectiveness. Subsequent studies should delve into ChatGPT's enduring influence on L2 academic writing and its potential role in boosting student motivation and evaluative judgement on feedback.

In conclusion, the integration of ChatGPT into L2 writing pedagogy holds significant potential, enhancing English proficiency and guiding students in identifying and rectifying errors through active interaction with the tool. While AWE tools provide invaluable support to L2 learners and reduce educators' evaluation load, their inherent limitations should be acknowledged. Complementing these tools with personalized teacher feedback can lead to a holistic assessment of students' writing progress. It's imperative to note that the primary objective of this paper isn't to substitute traditional assessment methods with AI, but rather to advocate for a harmonious integration of the two. Academic writing pedagogy rests on melding pioneering technology with proven teaching methodologies, upholding the primacy of academic integrity.

## References

- [1] Zhang, Z. V. (2020). *The Interplay between Students and AWE: Unpacking Feedback Reception and Perception*. *Assessing Writing*, 43, 100439.
- [2] Babak, K. (2019). *Scrutinizing the Limitations of AWE Tools: A Study on Grammarly*.
- [3] Ferster, B., Hammond, T. C., Alexander, R. C., & Lyman, H. (2012). *The Role of Automated Systems in Enhancing Student Writing*. *Journal of Interactive Learning Research*, 23(1), 81-99.
- [4] Shadiev, R., & Feng, Y. (2023). *Advancements in AWE: A Study on the Evolution of Feedback Tools*. *Interactive Learning Environments*.
- [5] Kasneci, E. et al. (2023). *ChatGPT for good? On opportunities and challenges of large language models for education*. *Learning and Individual Differences*, 103, 102274.
- [6] Blair, A., Curtis, S., Goodwin, M., & Shields, S. (2013). *What feedback do students want?* *Politics*, 33(1), 66-79.
- [7] Atlas, S. (2023). *ChatGPT for higher education and professional development: A guide to conversational AI*. Available at: [https://digitalcommons.uri.edu/cba\\_facpubs/548](https://digitalcommons.uri.edu/cba_facpubs/548). Accessed: July 23, 2023.
- [8] Khoshnevisan, B. (2019). *The affordances and constraints of automatic writing evaluation (AWE) tools: A case for Grammarly*. *ARTESOL EFL Journal*, 2(2), 12-25.
- [9] John, P., & Woll, N. (2020). *Using grammar checkers in an ESL context: An investigation of automatic corrective feedback*. *CALICO Journal*, 37(2), 169-192.
- [10] El Ebyary, K., & Windeatt, S. (2010). *The impact of computer-based feedback on students' written work*. *International Journal of English Studies*, 10(2), 121-142.

- [11] Singh, D., Meyer, A., & Young, J. (2017). *The impact of peer feedback on students' written assignment*. *Education Tech Research Dev*, 65, 203-226.
- [12] Newmann, F. M. (1990). *Higher order thinking in teaching social studies: A rationale for the assessment of classroom thoughtfulness*. *Journal of curriculum studies*, 22(1), 41-56.
- [13] Hou, Y. (2020). *Implications of AES system of Pigai for self-regulated learning*. *Theory and Practice in Language Studies*, 10(3), 261-268.
- [14] Cheville, J. (2004). *Automated scoring technologies and the rising influence of error*. *The English Journal*, 93(4), 47-52.
- [15] About Us. Grammarly. (n.d.). Retrieved April 1, 2023, from <https://www.grammarly.com/about>
- [16] Lim, H. & Kahng, J. (2012). *Review of Criterion for English Language Learning*. *Language Learning & Technology*, 16(2), 38-45. <http://dx.doi.org/10125/44285>
- [17] Dergaa, I., Chamari, K., Zmijewski, P., & Saad, H. B. (2023). *From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing*. *Biology of Sport*, 40(2), 615-622.
- [18] Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y. S., Gašević, D., & Chen, G. (2023, July). *Can large language models provide feedback to students? A case study on ChatGPT*. In *2023 IEEE International Conference on Advanced Learning Technologies (ICALT)* (pp. 323-325). IEEE.
- [19] Perkins, M. (2023). *Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond*. *Journal of University Teaching & Learning Practice*, 20(2), 07.
- [20] Zohery, M. (2023). *ChatGPT in Academic Writing and Publishing: A Comprehensive Guide*. In *Artificial Intelligence in Academia, Research and Science: ChatGPT as a Case Study*. (Frist Edition). Achtago Publishing. <https://doi.org/10.5281/zenodo.7803703>
- [21] Fitria, T. N. (2023, March). *Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay*. In *ELT Forum: Journal of English Language Teaching* (Vol. 12, No. 1, pp. 44-58).
- [22] Susnjak, T. (2022). *ChatGPT: The end of online exam integrity?* *arXiv preprint arXiv:2212.09292*.
- [23] Haque, M. U., Dharmadasa, I., Sworna, Z. T., Rajapakse, R. N., & Ahmad, H. (2022). *"I think this is the most disruptive technology": Exploring sentiments of ChatGPT early adopters using Twitter data*. *arXiv*.