Proceedings of ICLLCD 2026 Symposium: Using Visual Arts to Enrich History Understanding
DOI: 10.54254/2753-7064/2026.HT32003

Pedagogical Value and Limitations of Automated Writing Evaluation in English as a Second Language Writing Instruction

Ziling Huang

School of Foreign Languages and Cultures, Nanjing Normal University, Nanjing, China
05231728@njnu.edu.cn

Abstract. With the advancement of artificial intelligence and natural language processing technologies, the automated writing evaluation system (AWE) has gradually been integrated into English as a Second Language (ESL) writing instruction and has gained potential value for enhancing feedback efficiency and supporting writing revision. However, there are still some disputes over their instructional effectiveness and the boundaries of their application. This study analyzes the pedagogical value and limitations of AWE in ESL writing instruction through reviews of empirical studies, systematic reviews and meta-analyses published in the past 20 years. The findings show that AWE has a relatively stable positive effect on improving surface-level linguistic features such as grammar, spelling, and punctuation, and can sustain several rounds of modification and to a certain degree, improve learners' motivation and autonomy in writing. However, it is still unknown whether it is valid and stable enough to assess higher-order writing abilities such as writing content development, argumentation logic and appropriateness. This study argues that AWE should be viewed as a supplementary tool to provide formative assessments, functioning effectively in instructional models and human–AI collaboration. The findings of this study may provide pedagogical implications for the rational integration of AWE tools into ESL writing classrooms, offering directions for future research in this field.

Keywords: Automated Writing Evaluation (AWE), ESL writing instruction, formative assessment

1. Introduction

1.1. Research background

In a globalized environment, English is now an indispensable medium for international communication, academic research, and career development. Among all the language skills, English writing skill is not only an important way for academic communication and knowledge building, but also an index used in higher education and the international academic evaluation system to test learners' language proficiency and academic literacy. For learners of English as a Second Language

(ESL), writing is both a goal of language learning and one of the most challenging skills to master [1].

A large number of researches have shown that ESL learners often run into several difficulties in English writing, including grammar and vocabulary errors, insufficient discourse organization skills, weak argumentative logic, and high levels of writing anxiety [2,3]. These difficulties stem not only from limited mastery of linguistic forms but also from learners' unclear understanding of academic writing conventions, argumentative forms, and text coherence. In the realm of writing pedagogy, individualized feedback is usually considered an important contributor to students' writing development. However, in a real instructional situation, what should be noticed is the limitation of time, the large number of students and a heavy workload on grading. It is especially true for large-scale instruction and examination-oriented courses. Therefore, it is difficult for teachers to frequently, quickly and specifically instruct all learners in writing [4]. Against this backdrop, how to improve the efficiency of writing instruction without compromising on feedback quality has become a practical problem for ESL writing pedagogy.

With natural language processing and artificial intelligence technologies developing, automated writing evaluation (AWE) systems have been introduced into the ESL writing instruction field. As for instructional purposes, AWE systems facilitate timely and more accessible writing feedback, as well as continuous help for students in revising their writing. The systems can automatically analyze the text written by the students and give scores and feedback and so they can be seen as a way to relieve teachers' burden but still support writing development [4,5].

1.2. Research questions

Although the application of AWE in ESL writing instruction has become increasingly widespread, its educational value in writing outcomes, feedback effectiveness and instructional adaptability remains highly contested. On one hand, AWE can help improve writing accuracy, facilitate multiple rounds of revision, and support formative assessment. It has a strong, stable effect on enhancing surface-level language skills, such as grammar, spelling, and punctuation [6,7]. This positive impact is particularly evident in the consistent improvement of mechanical accuracy across multiple studies [8,9]. On the other hand, some studies have questioned AWE in fostering higher-order writing abilities, as it can provide little support for content development, argumentative logic, and contextual appropriateness. In addition, sometimes AWE may also make wrong evaluations, over-correct or even have negative consequences on learners' writing strategies and autonomy [4,10,11]. In particular, after the introduction of generative artificial intelligence into writing assessments, issues related to assessment validity, feedback stability, and the division of labor between humans and machines have further intensified scholarly debate [5].

In light of these divergences, it is necessary to examine the pedagogical value and limitations of AWE in ESL writing instruction to gain a more comprehensive understanding of its instructional potential.

1.3. Research significance

Based on the aforementioned background and identified issues, this study aims to review the current state of AWE applications and primary results concerning AWE in ESL writing instruction by examining relevant empirical studies, systematic reviews, and meta-analyses. It especially focuses on analyzing both the positive and negative effects of AWE on writing outcomes. Theoretically, it clarifies the divergences in research on the instructional effects of AWE and provides clearer guidance for understanding when and how AWE is effective. Practically speaking, it provides references for ESL writing teachers to make rational use of AWE tools in classrooms and improve feedback design, while also indicating potential directions for future research.

1.4. Paper structure

This paper contains five sections. Section One is the introduction where it shows the research background, research questions, research objectives, and the structure of the paper. Section Two is about the definition of AWE, the development of AWE, and the types of feedback it can offer. Section Three focuses on analyzing the pedagogical strengths and limitations of AWE in ESL writing instruction. Section Four builds on the preceding analysis to conduct further discussion, giving all the discoveries a big picture spin and thinking about how AWE acts and works when teachers use it. Section Five concludes the paper with a summary of the main findings, pedagogical implications, and research limitations, as well as directions for future studies.

2. Definition, developmental background, and feedback types of AWE

2.1. Definition and developmental background of AWE

2.1.1. Definition of AWE

Automated Writing Evaluation (AWE) refers to a system with evaluation functions using computer programs and natural language representations to automatically score and/or give feedback on a learner's writing text [4]. From the perspective of technological progress, AWE systems are divided into two types:

The first type consists of traditional AWE systems which mostly use rule-based or feature-based methods and study texts using set linguistic rules or statistical characteristics. These systems show improvement on stable performance for spelling, grammar, and mechanical errors, but the ability to handle discourse structure and the semantic level is limited [4].

The second type is AI-powered AWE systems, with generative systems based on machine learning and large language models (LLM). They can manage some complicated patterns to a certain extent and provide overall judgments or natural language feed, which gives a deeper level of information beyond just some basic score and narrows the scope of areas to be focused, thus providing more opportunities for different expressions [5,10]. (see Table 1).

Table 1. Differences between traditional AWE systems and ai-driven AWE systems

Dimension	Traditional AWE Systems	AI-Powered AWE Systems
Underlying approach	Rule-based or feature-based methods relying on predefined linguistic rules or statistical features	Machine learning–based approaches, including generative models and large language models (LLMs)
Primary analytical focus	Surface-level linguistic features such as spelling, grammar, and mechanical accuracy	Broader linguistic patterns, including discourse-level and holistic text features
Feedback type	Error detection and correction, often presented as explicit flags or rule-based suggestions	Holistic evaluations and natural language feedback, often expressed in explanatory or advisory forms
Strengths	High stability and consistency in identifying spelling, grammatical, and mechanical errors	Greater flexibility in handling complex language patterns and generating integrated feedback
Limitations	Limited capacity to process discourse structure, semantics, and higher-order writing features	Potential variability in feedback reliability and validity, especially at higher-order levels
Pedagogical role	Effective for improving surface-level language accuracy	Potential to support broader revision and reflection when guided by teachers

2.1.2. Developmental background of AWE

Early studies considered AWE mainly as an AI writing and scoring tool. Its purpose was to improve the writing assessment efficiency and consistency, making the assessment more objective, determining that the evaluation mainly included quantitative elements [4,12]. In this period, AWE was used mainly for high-stakes examinations and large-scale standardized testing contexts. Its evaluation criteria is depended on measurable linguistic traits including word quantity, sentence extent, grammatical correctness and mechanical mistakes [9].

With the emergence of the formative assessment principle in second language writing research, the function of AWE has gradually transformed from an automated scoring tool that is result-focused to an automated feedback system that fosters the process of writing and multiple rounds of revision in the classroom. Researchers started to stress that AWE would provide immediate feedback to the learners in the classrooms as supportive for the writing processes and revision rather than creating a terminal score. At present, most studies conceptualize AWE mainly as a technology supporting the learning process of writing and facilitating revising behaviors rather than substituting teachers' judgment. Its core pedagogical value lies in providing formative feedback for learners while alleviating teachers' workload in writing assessment [3,5].

In recent years, with the rapid development of machine learning and large language models (LLMs), AWE has continued to grow in terms of depth and form of feedback creation. Firstly, statistical and corpus-based methods are employed in combination with deep learning to bring about intelligent systems that are able to identify complicated aspects of language utilization. Secondly, new generative artificial intelligence tools, such as ChatGPT, have begun to be used to generate holistic, discourse-level writing evaluations and suggestions, which means that AWE is moving away from being driven by rules and features toward being driven by meanings and generating content [5,10]. However, despite ongoing technological advances, controversies about whether AWE covers the appropriate construct, whether it can evaluate and assess students' writing accuracy reliably, and whether its algorithm can be flexible enough to serve various writers persist despite technological improvements, requiring more classroom applications with proper instructional designs and teacher moderations [9,13].

2.2. Types of feedback provided by AWE

Existing research generally agrees that AWE does not provide a single form of feedback but covers multiple levels of feedback. According to the level of writing targeted by the feedback, AWE feedback can be categorized into surface, structural, and content feedback.

2.2.1. Surface-level feedback: spelling, grammar, and punctuation

At present, surface-level feedback is the most mature form of AWE feedback, which has the strongest supportive evidence among all forms. This is because sentence-level feedback mainly judges error proneness from language features like spelling, grammar, and punctuation, which show a high degree of uniqueness and rule sturdiness, making it possible for errors to be identified using rules or statistical methods. Surface-level feedback refers mainly to the errors in formal language use, such as spelling, grammar, punctuation, and vocabulary choices. Many experiments have shown that AWE systems possess high stability and accuracy in recognizing and annotating these kinds of errors, and the resulting surface-level errors are thus greatly reduced in learners' texts [6,7].

Second language writing accuracy improvements caused by surface-level feedback exhibit rather stable positive impacts on the condition that there are multiple chances of revising materials, that the feedback only targets mistakes involving grammar or mechanics, and that the task that writers complete clearly specifies a purpose [3,13]. For example, commonly used AWE tools such as Pigai, Grammarly, and Criterion can help learners identify issues like subjects and verbs that agree, tenses coincide, correct spellings and punctuations, thereby aiding writing improvements for lacking clarity and texts that have poor arrangement and organization [7,13].

2.2.2. Structural feedback: coherence and paragraph organization

Compared with surface-level feedback, the capacity of AWE for structural-level feedback has not been well-implemented, manifested as feedback accuracy and explanation level of paragraph organizational or discourse coherence failing to be fully met. Some AWE systems can give prompts about paragraph structure, sentence linkage, or discourse coherence on a macro-level like informing users whether a certain paragraph lacks a topic sentence, transitions are vague or the structural elements are unevenly distributed [14].

Structural feedback in this case is provided in a diagnostic way, so it is not as precise as or as detailed as teacher's feedback. This suggests that structural feedback cannot fully replace teacher expertise to evaluate how well students organize their discourse and rhetoric [5,9]. In particular, AWE often relies on formal features to infer coherence in complex writing tasks rather than understanding rhetorical intentions and information organization at the discourse level [4]. As a result, Thus, structural feedback, although it provides students with some cue words in these hard parts, plays mainly as a supplementary function in students' discourse organization rather than an independent instructional scaffold because it requires pedagogical mediations from teachers to instructively reinforce students' focusing on certain discourse contents that are sometimes also rather implicit under students' awareness [3].

2.2.3. Content feedback: logicality and argument support

Content feedback is regarded as the most disputed type of feedback supplied by the AWE currently, whereas higher-order writing abilities are the capacities involved with accomplishing complex tasks, generating ideas that are structured coherently, arguments that are constructed and evaluated cogently, or producing expressions that are targeted and appropriate given specific contexts and purposes. Higher-order writing feedback requires technical and theoretical challenges for both evaluators and technologies: judging argumentative logic, being precise about claims, providing strong evidence to support arguments, and making sense of various contexts/ fields/ disciplines. All of these depend on a deep-seated knowledge about context and discipline and evaluation on the basis of rhetoric [4].

Traditional AWE systems can barely supply content feedback and may even generate misjudgments or change the original intended meaning of the sentences [9,13]. Although generative AI has promising results regarding holistic evaluation and comments, there is a lack of sufficient stability and validity of content feedback. This is especially clear for writing tasks with integrated sources or academic argumentation, when AWE systems cannot make reliable evaluations on arguments' sufficiency or reasoning's rigor [5,10].

Subsequently, AWE can only offer heuristic advice or a general overview of what the content communicates, but it does not serve as an authoritative evaluation system to assess advanced writing abilities. The pedagogical value of the feedback based on content analysis provided by AWE highly relies on the teacher's selection, modification and use of the automatically generated suggestions, and helping students distinguish real constructive comments from irrelevant remarks, and guiding them towards higher-order writing revision goals [9,13].

3. Pedagogical value and limitations of AWE in ESL writing instruction

3.1. Pedagogical value of AWE in ESL writing instruction

3.1.1. Direct improvement of language and text quality

The major outcome from current studies seems to be that AWE has the strongest consistent facilitative influence on the surface aspects of language, such as grammar, spelling, punctuation, and certain types of lexical usage. It was found that learners could reduce lower-order mistakes significantly after many iterations of AWE, which would raise the text's comprehensibility and its adherence to conventions [6,7,13]. Additionally, most automated writing tools produce observable gains primarily with regard to spelling, grammar, and mechanical accuracy but not much feedback about structure, context or rhetoric. This pattern further corroborates the pedagogical value of AWE, which is most prominent in enhancing surface-level language quality [5].

Writing quality has aspects extending beyond surface-level linguistic features such as word choice and grammatical accuracy, namely those related to content development, argumentation logic, and discourse construction. Therefore, it is crucial to understand that this "improvement effect" pertains primarily to linguistic form and does not manifest in the full scope of writing quality enhancement. Meanwhile, many investigations also pointed out that the percentage of helpful feedback focusing on structure or content was far less than that of low-level feedback [5]. Such a difference is considered one of the key boundaries for understanding the pedagogical value of the AWE.

3.1.2. Promotion of affective factors and motivation

In the affective and motivational aspects, the pedagogical function of AWE mainly manifests as a cycle of "submission–feedback–revision–progress," in which the immediate feedback loop and visual evaluation indications like scores, levels, and dimensional reports can heighten learners' intention of revising to carry out the learning motivation [5,14]. Online automated feedback encourages students to engage in revision tasks, such as substitute, delete, reorganize, and expand and helps them adjust their intentions during the writing process, thus some learners will expend more time on revising an article with multiple drafts and a new one at last [15-17].

Apart from motivation, the non-interpersonal character of AWE feedback can relieve pressures on certain learners in face-to-face correction situations; yet the level of relief may differ across learner groups based on their different levels of writing anxiety, their degree of trust in feedback, and other context-specific variables. Previous empirical research has indicated that writing anxiety can affect learners' acceptance of AI-based evaluation tools, even if these particular tools can lower students' anxiety while being evaluated [5]. For example, anxiety can impact the ease of using an AWE tool (e. g., ChatGPT), and therefore learner attitudes about these tools can be explained through levels of perceived anxiety [18]. Recent meta-analyses reveal both anxiety and motivation to be correlated with each other and to impact writing competencies. These analyses have explicitly noted inconsistent findings regarding whether AWE reduces anxiety or enhances students' motivation.

Therefore, a more cautious conclusion is that AWE may offer potential affective benefits, but these benefits are highly dependent on implementation conditions and learner differences [8].

3.1.3. Enhancement of learner autonomy and writing agencytructural feedback: coherence and paragraph organization

In this study, learner autonomy refers to the ability to engage in self-monitoring, preliminary diagnosis, and revision decision-making during the writing process, which is closely related to self-regulated learning and feedback literacy.

From the view of the writing process, AWE can give more timely and accessible feedback than teachers' feedback. Therefore, part of the function of teachers in traditional classrooms is replaced, partly solving the problem that the frequency of the learners' receiving feedback is low, so learners could perform self-check and self-revision even without the help of their teachers [6,14]. This will improve independence and self-regulation during writing. Teacher interview articles, classroom observation studies show that the students who get the help of the AWE system will revise their work by themselves first. As time goes by, the learners might have some self-monitoring habits. And this type of process-oriented change could be regarded as an improvement of learner autonomy [14].

At the same time, learners show different trust tendencies in feedback at different levels. They usually trust grammatical and mechanical feedback more and are wary of content-level feedback. Select adoption of feedback shows emerging development of learners' critical engagement with feedback [5].

3.1.4. Support for instructional systems and classroom practice

At the instructional system level, the value of AWE is often summarized as "workload reduction and redistribution." By taking over some of the tasks in lower-level error correction and early diagnostics, AWE enables teachers to reallocate time and instructional resources and focus more on higher-order writing guidance, individualized support, and classroom interaction, as well as higher-order writing skills development. These skills include argumentative logic, content development, and genre awareness. Moreover, teachers can provide more differentiated support and individualized tutoring [4,14]. Several studies have reported positive or superior outcomes for blended feedback models that combine AWE and teacher feedback. Therefore, it would be better to classify AWE as a supplement rather than a replacement for teacher feedback [5].

3.2. Limitations of AWE in ESL writing instruction

3.2.1. Limitations of technical and algorithmic capacity

Whether it is a traditional or AI-powered AWE system, they still have limitations in content development, argumentation logic and context comprehension [4,9]. Key studies point out that most AWE systems can only recognize the surface-level features of language instead of making judgments regarding the content significance, logicality and organization, and their high-level feedback cannot compare to teachers' professional judgment [4]. Research on the use of ChatGPT for writing evaluation similarly suggests that in integrated writing tasks, human raters can draw on source texts and contextual understanding to infer the relationship between content and the materials. In contrast, ChatGPT demonstrated limited performance in understanding content relevance and argumentative nuances. Researchers have cautioned that such systems exhibit weaknesses in nuanced argumentation and contextual interpretation [19]. Moreover, most automated writing tools cannot tell writers which aspects of structure, context, and rhetoric need to be improved, thus lack of feedback leads to the fact that students can only repeatedly correct surface-level issues instead of improving argumentative strength [5].

3.2.2. Problems in feedback design

Another contributing factor to the deleterious effects of AWE originates from how such AWE presents itself to the student, as well as the extent to which it explains causes. Empirical evidence has shown that these factors may lead to over-correction and require a lot of cognitive effort from students. Given this situation, learners may undertake mechanized revisions rather than comprehend which constricts long-term learning transfer [6]. A related phenomenon is that some students regard AWE as unclear and difficult to understand, which may sometimes force them to comply entirely with computer-generated recommendations. These findings suggest that an increased amount of feedback is not necessarily more effective. Indeed, for feedback to be more useful, it must increase processing capacity rather than just increasing its volume [5].

In generative AI contexts, researchers have reported that AI feedback may only inform users what they should change but does not state why. This pattern decreases rule internalization and deeper studying [10]. In other words, without sufficient meta-linguistic explanations, AWE does not lead to sustainable language development despite a higher short-term revision efficiency [6,10].

3.2.3. Biases in learners' usage strategies

The third kind of risk comes from how learners use AWE tools. When learners regard AWE as an authoritative error corrector, writing becomes a matter of waiting for the system to find errors and revise accordingly. This approach will affect a learner's ability of self-monitoring and problem-solve [4]. Regarding the adoption of feedback, learners showed selective acceptance of AWE feedback. They'll generally prefer the grammatical/mechanical over content-level suggestions. Though this is selective, it could be due to the growing criticism of feedback and indicates the effort of revision will focus on surface-level features. As a result, students may ignore more important higher-order revisions needed to improve writing quality [5].

3.2.4. Constraints of external conditions and usage contexts

Finally, the actual effectiveness of AWE is limited by external conditions, such as network stability, device accessibility, and platform availability. Classroom-based studies indicate that technical barriers directly affect the learner's user experience. They will decrease the sustained interest and adoption of AWE tools [13]. According to new meta-analysis, research results vary by regions. Apart from differences in instructional implementation, different instructional results may also be caused by the differences in resources and usage environment [8].

The literature indicates it is clear that, in terms of how AWE affects ESL writing results, there is a definite hierarchy. There is the most evidence that there have been improvements in surface-level language quality as well as support for process-oriented revision. In contrast, there are still structural restrictions in the development of higher-order writing ability, feedback explanation depth, and contextual understanding [4,5,19]. Therefore, it is more feasible for educators to apply AWE in their instructional materials, situating it into a formative assessment framework where it collaborates with teachers to address deficiencies for maximizing positive impacts on teaching and learning [5].

4. Conclusion

According to the related systematic reviews, meta-analyses, and multiple empirical studies, this study explores the effects of Automated Writing Evaluation (AWE) on ESL learners' writing results. The comprehensive results reveal that AWE does have a consistent and important effect on improving students' writing effects, which is crucial for the enhancement of surface-level linguistic features, such as grammar, spelling, and punctuation. Likewise, all existing research clearly shows that AWE is a very limited tool in higher-order abilities required in writing, such as developing ideas, logical arguments and context. The validity and stability feedback of AWE were unable to fully substitute teachers' professional evaluations. This hierarchical asymmetry is a key to understanding the pedagogical effects of AWE.

From a pedagogical perspective, the findings of this study support a human–AI collaborative approach to writing instruction. The appropriate positioning of AWE is as a supplementary tool in a formative assessment framework depended on teachers' selection, interpretation, and remediation of feedback content, as well as explicit guidance on learners' modes of use. From another perspective, AWE can be responsible for low-level error correction and initial diagnoses so as to offer learning services to students on a continuous and immediate basis at the process level. However, the mediating role of teachers, in charge of higher-order writing guidance, feedback interpretation, and learning strategy teaching, cannot be ignored. The realization of beneficial and enduring benefits from AWE depends on such a framework constituted of teacher oversight and instructional design.

Even though there is a rather complete evidence base drawn from synthesizing past research, its conclusion is limited by existing research. First, current empirical studies of AWE focus only on short-term writing achievement or singular instructional interventions. Long-term investigations into long-term effects, such as writing skill transfer and the development of autonomous writing ability, remain insufficient. Second, the effectiveness of AWE is highly context-dependent, yet cross-cultural and cross-context comparative studies remain scarce. Moreover, little has been explored regarding whether AWE is feasible across different types of schools, different linguistic backgrounds, or different technological conditions. Future research should use longer time frames, the combination of different methods, and classroom-based studies to learn more about the long-term ways that AWE makes a difference in real classrooms.

In summary, AWE is a significant result of progress in educational technology and has shown considerable promise in ESL writing education in recent years. However, its value lies not in replacing teachers but in reshaping the approaches to writing assessment and feedback. Only through the continuous advancement of empirical research and the rational application of AWE in pedagogical practice can it truly make a difference in the development of second-language writing. In future research and practice, researchers and teachers should strengthen their cooperation to build closer connections among theory, technology, and classroom practice to promote the sustainable development of AWE in the field of language education.

References

[1] Kurt, G., Atay, D., & Öztürk, H. A. (2022). Student engagement in K12 online education during the pandemic: The case of Turkey. Journal of Research on Technology in Education, 54(sup1), S31–S47.

[2] Evans, S., & Green, C. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. Journal of English for Academic Purposes, 6(1), 3–17.

[3] Fan, N., & Ma, Y. (2022). The effects of automated writing evaluation (AWE) feedback on students' English writing quality: A systematic literature review. Language Teaching Research Quarterly, 28, 53–73.

DOI: 10.54254/2753-7064/2026.HT32003

[4] Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. Language Teaching Research, 10(2), 1–24.

[5] Liu, W. (2024). A systematic review of automated writing evaluation feedback: Validity, effects and students' engagement. Language Teaching Research Quarterly, 45, 86–105.

[6] Barrot, J. S. (2021). Using automated written corrective feedback in the writing classroom: Effects on L2 writing accuracy. Computer-Assisted Language Learning. Advance online publication.

[7] Zhang, X. (2022). Improving Pigai as an automated writing evaluation system: Considerations for refinement. Frontiers in Psychology, 13, Article 795725.

[8] Huang, X., Xu, W., Li, F., & Yu, Z. (2024). A meta-analysis of effects of automated writing evaluation on anxiety, motivation, and second language writing skills. The Asia-Pacific Education Researcher, 33, 957–976.

[9] Shi, H., & Aryadoust, V. (2022). A systematic review of automated writing evaluation systems. Education and Information Technologies. Advance online publication.

[10] Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. International Journal of Educational Technology in Higher Education, 20(57).

[11] Shermis, M. D., & Wilson, J. (2024). Introduction to automated essay evaluation. In The Routledge international handbook of automated essay evaluation (pp. 3–22). Routledge.

[12] Tang, J., & Rich, C. S. (2017). Automated writing evaluation in an EFL setting: Lessons from China. JALT CALL Journal, 13(2), 117–143.

[13] Rahman, N. A., Zulkornain, L. H., Che Mat, A., & Kustati, M. (2023). Assessing writing abilities using AI-powered writing evaluations. Journal of Asian Behavioural Studies, 8(24), 1–17.

[14] Wilson, J., Ahrendt, C., Fudge, E. A., Raiche, A., Beard, G., & MacArthur, C. A. (2021). Elementary teachers' perceptions of automated feedback and automated scoring: Transforming the teaching and learning of writing using automated writing evaluation. Computers & Education, 168, 104208.

[15] Zhang, Z. V. (2017). Student engagement with computer-generated feedback: A case study. ELT Journal, 71(3), 317–328.

[16] Roscoe, R. D., Wilson, J., Johnson, A. C., & Mayra, C. R. (2017). Presentation, expectations, and experience: Sources of student perceptions of automated writing evaluation. Computers in Human Behavior, 70, 207–221. https: //doi.org/10.1016/j.chb.2016.12.076

[17] Sung, Y. T., Liao, C. N., Chang, T. H., Chen, C. L., & Chang, K. E. (2016). The effect of online summary assessment and feedback system on the summary writing on 6th graders: The LSA-based technique. Computers & Education, 95, 1–18.

[18] Kim, H., Baghestani, S., Yin, S., Karatay, Y., Kurt, S., Beck, J., & Karatay, L. (2024). ChatGPT for writing evaluation: Examining the accuracy and reliability of AI-generated scores compared to human raters. In C. A. Chapelle, G. H. Beckett, & J. Ranalli (Eds.), Exploring artificial intelligence in applied linguistics (pp. 73–95). Iowa State University Digital Press.

[19] Liu, Z. (2024). Second language writing anxiety and ChatGPT adoption as an automated writing evaluation tool. Journal of Applied Research in Higher Education. Advance online publication.